

AA003705 331 bp mRNA
mg1a02.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
clone IMAGE:437450 5', mRNA sequence.
AA003705
AA003705.1 GI:1447205
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 331)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geissel, S., Kucaba, T., Lacy, M., Le M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Threising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R., and
Waterston, P.

REFERENCE
AUTHORS

TITLE	JOURNAL	COMMENT
1. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See page 100
2. The Impact of Globalization on the Labor Market	Journal of International Economics	See page 100
3. The Effect of Trade Liberalization on the Environment	Journal of Environmental Economics	See page 100
4. The Role of the State in the Development of the Environment	Journal of Environmental Economics	See page 100
5. The Impact of Globalization on the Environment	Journal of International Economics	See page 100
6. The Effect of Trade Liberalization on the Environment	Journal of Environmental Economics	See page 100
7. The Role of the State in the Development of the Environment	Journal of Environmental Economics	See page 100
8. The Impact of Globalization on the Environment	Journal of International Economics	See page 100
9. The Effect of Trade Liberalization on the Environment	Journal of Environmental Economics	See page 100
10. The Role of the State in the Development of the Environment	Journal of Environmental Economics	See page 100

Contact: Marra M/Mouse EST Project
WashU-HMMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:262786
Seq primer: ETPRimer.

SEQ ID NO: 55

Tue Apr 30 14:18:38 2002

us-09-248-]

FEATURES source

```

Location/Qualifiers
1. 331
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:437450"
/clone.lib="Soares mouse embryo NbME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/note="Host: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTACCACTGACCTGAAGTGGAGCGCGCGGAAATTTTTTTTTTTT
T 3'] on an equal amount of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M. Fatima Bonaldo. "

```

BASE COUNT	72 a	65 c	73 g	121 t
ORIGIN				

```

Query Match          2.5%; Score 23; DB 10; Length 331;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

911	tatctgcaaaaaaaaaa	933
298	TATCTGCCAAAAAAA	320

XX 31-OCT-1996; 96US-0029960.
 PR (HUMA-) HUMAN GENOME SCI INC.
 XX
 PA Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
 PI Kunsch CA, Rosen CA;
 XX WPI; 1998-27225/24.
 DR
 XX Computer-readable medium with recorded Streptococcus pneumoniae
 PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus
 PT pneumoniae.
 XX
 PS Claim 1; Page 608-617; 1409pp; English.
 XX
 CC The present invention describes a computer readable medium which has
 CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
 CC recorded on it, or a representative fragment or a sequence at least 95%
 CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
 CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
 CC Streptococcus pneumoniae. The present invention also describes an
 CC isolated nucleic acid molecule encoding a homologue of any of the
 CC fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the
 CC nucleic acid molecule is produced by a process comprising: (a) screening
 CC a genomic DNA library using as a probe a target sequence defined by any
 CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
 CC library which contain sequences that hybridize to the target sequence and
 CC isolating the nucleic acid molecules from the members; or (b) isolating
 CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
 CC molecules whose nucleotide sequence is homologous to amplification
 CC primers derived from the fragment of the S. pneumoniae genome to prime
 CC the amplification and isolating the amplified sequences. The computer
 CC readable medium can be used in a computer-based system for identifying
 CC fragments of the S. pneumoniae genome of commercial importance, or
 CC expression modulating fragments of the S. pneumoniae genome. Products
 CC from the present invention can be used in diagnosis kits and assays, and
 CC pharmaceutical compositions and vaccines for S. pneumoniae.
 XX
 SQ Sequence 16535 BP; 5001 A; 3914 C; 3102 G; 4518 T; 0 other;

Query Match 2.8%; Score 17; DB 19; Length 16535;
 Best Local Similarity 100.0%; Pred. No. 86;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 474 ttcaaggatgcccaagaa 490
 |||||
 Db 451 ttcaaggatgcccaagaa 467

SEQ.ID.NO:60

RESULT 43
 AAV52207
 ID AAV52207 standard; DNA; 16535 BP.
 XX
 AC AAV52207;
 XX
 DT 23-OCT-1998 (first entry)
 XX
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:74.
 XX
 KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 KW computer readable medium; vaccine; pharmaceutical composition; ds.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO9818931-A2.
 XX
 PD 07-MAY-1998.
 XX
 PF 30-OCT-1997; 97WO-US19588.

Tue Apr 30 14:18:55 2002

US-08-998-416-136

Query Match 5.0%; Score 68; DB 4; Length 703;
 Best Local Similarity 100.0%; Pred. No. 8.7e-18;
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1299 gccaccgcggtgagctccagcttttgcctttagtgagggtaattgcgcgttgcc 1358
 |||||
 DB 158 GCCACCGCGGTGGAGCTCCAGCTTTTGTCCCTTTAGTGAGGGTTAATTGCGCGCTTGC 217
 |||||

QY 1359 gtaatacat 1366
 |||||
 DB 218 GTAATCAT 225

SEQ ID NO: 61

RESULT 1
 US-08-998-416-136
 ; Sequence 136, Application US/08998416
 ; Patent No. 6239264
 ; GENERAL INFORMATION:
 ; APPLICANT: Philippsen, Peter
 ; APPLICANT: Pohlmann, Rainer
 ; APPLICANT: Steiner, Sabine
 ; APPLICANT: Mohr, Christine
 ; APPLICANT: Wendland, Jurgon
 ; APPLICANT: Knechtle, Philipp
 ; APPLICANT: Rebischung, Corinne
 ; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPPII
 ; TITLE OF INVENTION: AND USES THEREOF
 ; NUMBER OF SEQUENCES: 1152
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 6239264artis Corporation
 ; STREET: 3054 Cornwallis Road
 ; CITY: Research Triangle Park
 ; STATE: No. 6239264th Carolina
 ; COUNTRY: USA
 ; ZIP: 27709
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/998,416
 ; FILING DATE: 24-DEC-1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: CH 0016/97
 ; FILING DATE: 31-DEC-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Meigs, J. Timothy
 ; REGISTRATION NUMBER: 38,241
 ; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 919-541-8587
 ; TELEFAX: 919-541-8689
 ; INFORMATION FOR SEQ ID NO: 136:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 703 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; ORIGINAL SOURCE:
 ; ORGANISM: PAG1051UP

SEQ ID NO: 62

RESULT 30
US-08-578-649-4
; Sequence 4, Application US/08578649
; Patent No. 5770366
; GENERAL INFORMATION:
; APPLICANT: Ulrich Bogdan
; APPLICANT: Reinhard Buttner
; APPLICANT: Brigitte Kaluza
; TITLE OF INVENTION: MELANOMA-INHIBITING PROTEIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/578,649
; FILING DATE: 29-July-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 43 24 247.2
; FILING DATE: 20-July-1993
; ATTORNEY/AGENT INFORMATION:

NAME: Andrew L. Tiajolloff
REGISTRATION NUMBER: 31,575
REFERENCE/DOCKET NUMBER: BOER 1035-PFF/ALT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 581 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 110..499
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 110..178
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 179..499
US-08-578-649-4

Query Match 1.9%; Score 18; DB 1; Length 581;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 907 ggcaaaaaaaaaaaaaaa 924
|||||||
Db 563 GGCAAAAAAAAAAAAAAA 580

SEQ ID NO: 63

RESULT 17
 AAV55038
 ID AAV55038 standard; cDNA; 5232 BP.
 XX
 AC AAV55038;
 XX
 DT 13-NOV-1998 (first entry)
 XX
 DE Human XIAP coding sequence.
 XX
 KW Inhibitor of apoptosis protein; apoptosis enhancer; NAIP polypeptide;
 KW proliferative disease; IAP; therapy; cancer; human; XIAP protein; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 34..1527
 FT /*tag= a
 FT /product= XIAP
 XX
 PN WO9835693-A2.
 XX
 PD 20-AUG-1998.
 XX
 PF 13-FEB-1998; 98WO-IB00781.
 XX
 PR 13-FEB-1997; 97US-0800929.
 XX
 PA (UYOT-) UNIV OTTAWA.
 XX
 PI Baird S, Korneluk R, Liston P, Mackenzie AE, Pratt C;
 PI Tsang B;
 XX
 DR WPI; 1998-467164/40.
 DR P-PSDB; AAW69294.
 XX
 PT Inducing apoptosis in proliferative mammalian cells with inhibitor
 PT of IAP or NAIP polypeptide - also methods for prognosis based on
 PT presence of IAP and NAIP, specifically applied to cancers involving
 PT p53 mutations
 XX
 PS Claim 13; Fig 1; 147pp; English.
 XX
 CC This sequence encodes the human XIAP protein, which is a inhibitor of
 CC apoptosis protein (IAP), and can be used in the method of the invention.
 CC The method is for enhancing apoptosis in cells from a mammal with
 CC proliferative disease by treatment with a compound that inhibits
 CC biological activity of an IAP or NAIP polypeptide. The inhibitory
 CC compounds are used to treat proliferative diseases, specially cancers of
 CC ovary, breast, pancreas, lymph nodes, skin, blood, lung, brain, kidney,
 CC liver nasopharynx, thyroid, central nervous system, prostate, colon,
 CC rectum, cervix or endometrium, particularly to increase their sensitivity
 CC to chemotherapeutic agents. High levels of the IAP or NAIP proteins are
 CC detected in many cancers and are associated with poor prognosis,
 CC resistance to chemotherapeutic agents and mutations in p53 (it is
 CC suggested that wild-type p53 suppresses transcription of the IAP or NAIP
 CC genes). Transgenic animals are used for testing the effects of antisense
 CC oligonucleotides and for screening for the inhibitors.
 XX
 SQ Sequence 5232 BP; 1579 A; 861 C; 1062 G; 1728 T; 2 other;

Query Match 1.9%; Score 21; DB 19; Length 5232;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 agtttttctaataccaagaagg 479
 |||||
 Db 2498 agtttttctaataccaagaagg 2518

RESULT 18

SEQ ID NO 64

RESULT 7
AAQ59633/c
ID AAQ59633 standard; cDNA; 319 BP.
XX
AC AAQ59633;
XX
DT 16-MAR-1994 (first entry)
XX
DE Human brain Expressed Sequence Tag EST01493.
XX
KW Gene transcription product; genetic markers; tagging; in vivo;
KW transcription; mapping; locations; chromosomes; chromosomal; ss.
XX
OS Homo sapiens.
XX
PN W09316178-A.
XX
PD 19-AUG-1993.
XX
PF 12-FEB-1993; 93WO-US01294.
XX
PR 12-FEB-1992; 92US-0837195.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
XX
PI Adams MD, Moreno RF, Venter CJ;
XX
DR WPI; 1993-272882/34.
XX
PT Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in-vivo, facilitate tagging
XX of most human genes
PS Example 4; Page 207; 500pp; English.
XX
CC The Expressed Sequence Tag was isolated from a human brain cDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes
CC on chromosomes, for individual or forensic identification, for mapping
CC locations of disease-associated genes, for identification of tissue
CC type, and for prepn. of antisense sequences, probes and constructs.
CC EST01493 has a "poor" coding probability as evaluated using the
CC coding-region prediction program CRM. See also AAQ59041-Q61440.
XX
SQ Sequence 319 BP; 118 A; 42 C; 75 G; 82 T; 2 other;

Query Match 2.7%; Score 27; DB 14; Length 319;
Best Local Similarity 100.0%; Pred. NO. 0.12;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 341 cttattttcacttaacataatgacctcc 367
Db 90 CTTATTTCACCTTAACATAATGACCTCC 64

RESULT 8

© 2000 Blackwell Science Ltd *Journal of Internal Medicine* 247: 399–406

Query Match 55.5%; Score 319; DB 4; Length 654;
Best Local Similarity 100.0%; Pred. NO. 2.7e-129;
Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2002, 19:34:49 ; Search time 8020.64 Seconds
(without alignments)
1570.670 Million cell updates/sec

Title: US-09-248-178-60
Perfect score: 602
Sequence: 1 tgaagacgcgcgggtggag.....tcattgcggaggttcagac 602

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vi: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htgo_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Match Length DB ID Description

1	571	94.9	2168	9	BC011675	9	BC011675
2	571	94.9	2228	9	HSA005890		
3	571	94.9	2280	9	BC000972		
4	459.6	76.3	2179	10	AF229637		
5	459.6	76.3	2194	10	BC011195		
6	121.8	20.2	130788	9	AF235097		
c	7	101.2	15.8	30858	10	AF277994	
c	8	88.2	11.7	162389	2	AC097025	
c	9	84.2	14.0	3681	10	AF277991	
c	10	84.2	14.0	3739	10	AF277992	
c	11	77.5	12.9	1828	3	AY058359	
c	12	77.6	12.9	97259	2	AC014936	
c	13	77.6	12.9	181437	3	AC010751	
c	14	77.5	12.9	186002	3	AC006933	
c	15	77.6	12.9	280887	3	AE003526	
c	16	57.2	9.5	142277	2	AC016204	
c	17	56.4	9.4	215046	2	AC011767	
c	18	56.2	9.3	206976	2	AC105301	
c	19	56.2	9.3	221657	2	AC105305	
c	20	48.4	8.0	110000	2	LMFLCHR34_16	
c	21	48.2	8.0	229329	2	AC093352	
c	22	46.5	7.7	159971	2	AC091951	
c	23	46	7.5	184451	9	AC016768	
c	24	46	7.5	219521	2	AC099414	
c	25	45.8	7.6	211973	2	AC090656	
c	26	45.5	7.6	173294	2	AC105317	
c	27	45	7.5	125020	9	AF429315	
c	28	45	7.5	203984	2	AC067854	
c	29	44.4	7.4	183552	2	AC104807	
c	30	44	7.3	3720	14	S76368	
c	31	44	7.3	43658	14	HSV3PRGEN	
c	32	44	7.3	112930	14	HSGEN	
c	33	43.8	7.3	7218	6	I66494	
c	34	43.6	7.2	106320	8	AC002304	
c	35	43.5	7.2	109171	8	AC002328	
c	36	43.4	7.2	178022	2	AC018348	
c	37	43.4	7.2	179337	2	AC091304	
c	38	43.2	7.2	200019	2	AC023310	
c	39	43	7.1	2012	5	LFL289860	
c	40	43	7.1	2931	14	AF192756	
c	41	42.8	7.1	229706	2	AC095672	
c	42	42.4	7.0	3353	3	ACIFPA	
c	43	42.4	7.0	4513	9	HSU53786	
c	44	42.4	7.0	6457	9	HSU53786	
c	45	42.4	7.0	150244	2	AC040980	

ALIGNMENTS

RESULT 1

BC011675

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

BC011675 Homo sapiens, similar to JMI protein, clone MGC:15381
IMAGE:4299954, mRNA, complete cds.

BC011675
BC011675.1 GI:15079716
MGC.
human.

Homo Sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2168)
Direct Submission
Submitted (30-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: ncapbs-r@mail.nih.gov
Tissue Procurement: ATCC

LPDGTIANLAKLQLVVNSAQRVHILAGQWERHVPVLLAEYHILPKLQDCRELESSRRL
AEIOELHOSVRAAAEAEARRKEEVYKQLMSELETLPDVSRLAYTORILEIVGNIRKQK
FEFTIKILSDIRELOKEINSLSGKILDRITFAVTDIEIVFKDAKKDDAVRKAYKYLAALHEN
CSOLJOTIETDGTIMREVPDLEFEOIETELGKKTIILNLEKIREDYPAILOENAGLGRV
REA"

BASE COUNT 438 a 597 c 590 g 403 t

Query Match 94.9%; Score 571; DB 9; Length 2280;
Best Local Similarity 98.7%; Pred. No. 2,1e-121;
Matches 597; Conservative 0; Mismatches 5; Indels 3; Gaps 2;

QY 1 tgaagaagccgcggtgagctgctgcccgcgctgagactgcccaacctggcgaagctgacgc 60
DB 1240 TGAAGAAGCCGCGGTGAGCTGCTGCCGCGATGGGACTGCCAAGCTTGCACAGCTCCAGC 1299
QY 61 ttgtgtgagaataagtcgccagcaggtcatccacttgcgagctcaatgaagaagacacc 120
DB 1300 TTGTGTTGAGAAATAGTGCACAGCGGTCTACCTTGGCGGTGAGTGGGAGAGACACC 1359
QY 121 ggggtccca--tctctgtgagtcaccgcca-ctccgaagctgcagattacagagagctgg 177
DB 1360 GGGTCCCACTCTCGCTGAGTACCGCCACCCTCCGAAGCTGCAGAGATTCACAGAGCTGG 1419
QY 178 aatcttctcagcagctgagcagagatccagaagactgcacagagatctccggcggtctg 237
DB 1420 AATCTTCTCGACGGCTGGCAGAGATCCAAAGACTGCACAGAGTGTCCGGCGGCTGCTG 1479
QY 238 aagaagccgcaggaagaggggtctataagcagctgagtcgaagctggagactctac 297
DB 1480 AGAGGCGCCGAGGAAGAGGAGGTCTATAAGCAGCTGATGTACAGCTGGAGACTCTGC 1539
QY 298 ccagagatgtgtcccgctggtgctacaccagcgcatctcgagatcgtgggcaacatcc 357
DB 1540 CCAGAGATGTGTCCCGGCTGGCTACACCAGCGCATCTCTGAGATCGTGGCAACATCC 1599
QY 358 ggaacagaagaagagatcccaagatcttctgatacgaagagcttcaagaagaaa 417
DB 1600 GGAACGAGAAGGAAGAGATCACCAAGATCTTGTCGATACGGAAGAGCTTCAGAGAGAAA 1659
QY 418 tcaactccctatctgggaagctggaccggagctttcggtgactgagctgtgattca 477
DB 1660 TCAACTCCCTATCTGGGAAGCTGGACCGGACGCTTTCGGGTGACTGATGAGCTTGTCTCA 1719
QY 478 aggatgccagaagaagcagctgctgttcggaagcctataagtatctagctctctcagc 537
DB 1720 AGGATGCCAAGAAGACGATGCTGTTCGGAAGGCCCTATAAGTATCTAGTGTCTGTCAGC 1779
QY 538 aagaactgcagcagctcatccagaccatcgagagacagacagaccatcatccggagattc 597
DB 1780 AGAACTGCAGCCAGCTCATCCAGACCATCCAGACACAGGCACCATCATGCGGGAGGTTTC 1839
QY 598 gagac 502
DB 1840 GAGAC 1844

RESULT 3
LOCUS BC000972
DEFINITION Homo sapiens, JM1 protein, clone MGC:5137 IMAGE:3449051, mRNA,
complete cds.
ACCESSION BC000972
VERSION BC000972.2 GI:12803024
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2280)
AUTHORS Strausberg,R.
TITLE Direct Submission

JOURNAL

REMARK
COMMENT

Submitted (16-NOV-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
On Feb 14, 2001 this sequence version replaced gi:12654302.
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalon@bcm.tmc.edu
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRAK Plate: 3 Row: n Column: 6
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 3114815.

FEATURES
source

1..2280
/organism="Homo sapiens"
/db_xref="LocusID:28952"
/db_xref="Taxon:9606"
/clone="MGC:5137 IMAGE:3449051"
/tissue_type="Placenta, choriocarcinoma"
/clone_lib="NIH_MGC_10"
/lab_host="DH10B"
/note="Vector: pCHV-SPORT6"
118..2001
/codon_start=1
/product="JM1 protein"
/protein_id="AAH00972.1"
/db_xref="GI:12654303"

CDS

/translation="MEADRILIHSLRQAGTAVPDVOTLRAFTTELVEAVVRLRV
INPAGSLSPILLPLAMSAARFPAISLAQAMDGLGYPLELGYQNFLYPSEPLDRLL
FLAEPDPTDASEDADOPAGDSAILLRAIGSIRDOLALPWVPHLTPKLOHQSSAL
OKPFIHSLRVVPELSSRGCEPREFOASPLLPVTPQVOPVGRVSALEHHAQLCCQT
GRDRPDEDVHRTSRLPPOEDTRQORLOKOLTEHLRQSLGLGAPTOARDLELL
QWAGAKATGAPKGSFRTHSEKFTFPLEPQAOATQVSDVPATSRPEQVTRAAQOEL
ESLRQLEGVNRISIEVEADMKTLGVSEVQAESECHSKSLSTAEERQDLRRAVEL
LPDGTIANLAKLQLVVNSAQRVHILAGQWERHVPVLLAEYHILPKLQDCRELESSRRL
AEIOELHOSVRAAAEAEARRKEEVYKQLMSELETLPDVSRLAYTORILEIVGNIRKQK
FEFTIKILSDIRELOKEINSLSGKILDRITFAVTDIEIVFKDAKKDDAVRKAYKYLAALHEN
CSOLJOTIETDGTIMREVPDLEFEOIETELGKKTIILNLEKIREDYPAILOENAGLGRV
REA"

BASE COUNT 450 a 712 c 599 g 409 t

Query Match 94.9%; Score 571; DB 9; Length 2280;
Best Local Similarity 98.7%; Pred. No. 2,1e-121;
Matches 597; Conservative 0; Mismatches 5; Indels 3; Gaps 2;

QY 1 tgaagaagccgcggtgagctgctgcccgcgctgagactgcccaacctggcgaagctgacgc 60
DB 1271 TGAAGAAGCCGCGGTGAGCTGCTGCCGCGATGGGACTGCCAAGCTTGCACAGCTCCAGC 1330
QY 61 ttgtgtgagaataagtcgccagcaggtcatccacttgcgagctcaatgaagaagacacc 120
DB 1331 TTGTGTTGAGAAATAGTGCACAGCGGTCTACCTTGGCGGTGAGTGGGAGAGACACC 1390
QY 121 ggggtccca--tctctgtgagtcaccgcca-ctccgaagctgcagattacagagagctgg 177
DB 1391 GGGTCCCACTCTCTCGCTGAGTACCGCCACCCTCCGAAGCTGCAGAGATTCACAGAGACTGG 1450
QY 178 aatcttctcagcagctgagcagagatccagaagactgcacagagatgtccggcggtctg 237

Db 1451 AATCTCTCGAGCGCTGCGACAGATCCAAAGAACTGCACAGAGTGTCCGGCGGCTGCTG 1510
Oy 238 aagagcccgacagaaagagaggtctataagcagctcaatcagaaactagaaactctgc 297
Db 1511 AAGAGCCCGCAGGAAGAGAGGAGTCTATAGCAGCTGATGTACAGCTGGAGACTCTGC 1570
Oy 298 ccagagatgtatcccgctgagcctacacccagcgcatcctggagatcgtgggaacatcc 357
Db 1571 CCAGAGATGTGTCCCGGTGCGCTACACCCAGCGCATCTCTGGAGATCGTGGGTAACATCC 1630
Oy 358 ggaagcaagaagaagatcaacagatctgtctgatacagaagaagcttcagaagaaa 417
Db 1631 GGAACGAAGAAGAGATACACCAAGATCTTGTCTGATACGARGAGCTTCAGAAAGAAA 1690
Oy 418 tcaactcctatctgggaagctgagcggagcgtttgagtgagctatagagcttgtttca 477
Db 1691 TCAACTCCTATCTGGGAAGCTGGACCGAGCTTTGCGGTGACTGATGAGCTTGTGTTC 1750
Oy 478 aggatgcaagaagaagacatctgtctcgaaggccctataagatctagctctgcaag 537
Db 1751 AGGATGCCAAGAAGGACCAATGCTGTTCGGAAGGCCATAAGTATCTAGCTGCTGCACG 1810
Oy 538 agaactcgagcagctctccagaccatcgagaccacagacagcagcaccatcatcgaggaggtc 597
Db 1811 ACAACTCGAGCAGCTATCCAGACCATCGAGGACACAGCACCATCATCGGGAGGTTTC 1870
Oy 598 gagac 602
Db 1871 GAGAC 1875

RESULT 4
LOCUS AF229637 2179 bp mRNA linear ROD 22-JUN-2000
DEFINITION Mus musculus Dxm40e protein (Dxm40e) mRNA, complete cds.
ACCESSION AF229637
VERSION AF229637.1 GI:7673613
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE Means,G.D., Toy,D.Y., Baum,P.R. and Derry,J.M.J.
AUTHORS A transcript map of a 2-Mb BAC contig in the proximal portion of the mouse X chromosome and regional mapping of the scurfy mutation
TITLE Genomics 65 (3), 213-223 (2000)
JOURNAL 2031388
MEDLINE 10857745
PUBMED 2 (bases 1 to 2179)
REFERENCE Derry,J.M.J.
AUTHORS Direct Submission
TITLE Submitted (31-JAN-2000) Molecular Biology, Immunex Corporation, 51 University St., Seattle, WA 98101, USA
JOURNAL Location/Qualifiers
FEATURES
source 1..2179
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/map="between DXMit123 and DXMit161"
1..2179
/gene="Dxm40e"
66..1952
/gene="Dxm40e"
/note="DNA segment, Chr X, Immunex-40, expressed"
/codon_start=1
/product="Dxm40e protein"
/protein_id="AAF65951.1"
/db_xref="GI:7673614"
/translation="MEADRILIHSLFQAGTAVFPEVOTLPAFTLVEAVVPCLRV
INPDVGSLSHLPLPMSARFLMSLAQCHMDLGYPLEGYORFLYSEPDRLILL
FLAERLFDSASEDADQAGDSAIFLRAIGSQIRI:JUALPWVPLLRIPKVRIOGSL

QPFHSSRLVLPENSSGELNEFQASPLLLPAPTQVPOLOGRAASLEHHASQLCQHV
NRDCGDEDRVMSRVSDSRAPORLRLKOLIEHLRWSWGLGAPTQVRDLGEM
QTWGRAMTGVKPKSRFTSEKFTFHLEPQVQAQVADVPATSOLEODTRAQEOEL
ESLREOLASVNHNIIEVEADMTLGINLVQVETECRSQSELSVAEOQALRLKSRVLEI
LPDGAANLAKLOLVYVESSAORLIHLASOWEKHVRVPLLAERYHLRLQDRCLESSRL
AEIQELHSHVRAAAEAPRKVEVYKOLYSELETLPKYSRLAYTORILEIVGNIRKQK
EETIKLSDTIKELONEINSLSKLDRFTAVTDLYPKDAKKDDAVKAYKYLAAUHEN
CSOLIOITIEDITGIMHREVRLDEQIETEMGRKLTLSNLEKICEDYRALROENAGLLRV
REA"
BASE COUNT 492 a 612 c 629 g 446 t
ORIGIN
Query Match 76.3%: Score 459.6; DB 10; Length 2179;
Best Local Similarity 87.3%: Pred. No. 9.7e-96;
Matches 527; Conservative 0; Mismatches 74; Indels 3; Gaps 2;
Oy 1 tgaagaagccgcgctgagctgctgccatgagactggagctgccaacacctgccaagctgcagc 60
Db 1222 TGAAGACGACGACGGTGAATTTGCTGCTGATGGGGTGCACACCTCGCCAACTGTGCAGC 1281
Oy 61 ttctggttgagaatagtgcccagcgggtcactccacttggcgggtcagtggaagaagcacc 120
Db 1282 TTGTGGTGGAGAGCAGCGCTCAGAGGGCTCATCTAGCAAGCCAGTGGGAGAAACACC 1341
Oy 121 ggggtccca--tctctgtgagtaccqcca-ctccgaagctgcagagattgcagagaagctgg 177
Db 1342 GGGTCCCACTTCTTCTGAGTAGTACCGCCACCTCAGAGACTCCAGGACTGTAGGAGGTGG 1401
Oy 178 aatcttctcgagcgtgagcagagatccaaagactgcaccagagtgctcggggcgagctgctg 237
Db 1402 AATGCTGTCTGACGGCTGGCAGAAATCCAGGAGCTGCACACACAGCTGTTCGAGCAGCTGCAG 1451
Oy 238 aagagcccgacagaaagagagagctctataagcagctgagctgcagagctggagactctgc 297
Db 1452 AGGAGGCTCGCCGGAAGAGGAGGAGTCTATAGCAGCTGGTGTCTAGAGCTGTGAGACTCTGC 1521
Oy 298 ccagagatgtctcccgctggcctacacccagcgcatcctcggagatcgtgggcaacatcc 357
Db 1522 CAAAGAGATGTGTCCCGCTGCTTATCTCAGCGCATCTCTGGAGATTGTGGGCAACATCC 1581
Oy 358 gaaagcgaagaagaagatcaccagatcttctgtctgatacagaagagcttcagaagaaa 417
Db 1582 GGAAGCAGAAGAGAGATCACTAAGATCTTGTCCGACACAAAGAACTCAGAAAGGAAA 1641
Oy 418 tcaactcctatctgggaagctgagcggagcgtttgagtgagctgagcttgcttca 477
Db 1642 TCAACTCTCTCTCTGGGAAGCTCGACCGACATTTGCAGTCACTGATGAGCTGTGTGTCA 1701
Oy 478 aggatgcaagaagaagacatcgtctcgaaggcctataagatctagctctgctgcaag 537
Db 1702 AGGATGCTAAGAGGATGATGCTGTTCGAAAGGCTTACAGTACTAGCTGCCCTGCAATG 1761
Oy 538 aagactgagcagcagctcctccagaccatcgaggacacagggcaccatcagcggagagttc 597
Db 1762 AGAATTGCAGTCAGCTCATCCAGACCAATTGAGACACAGGCACTATCATAGGGAAGTTC 1821
Oy 598 gaga 501
Db 1822 GAGA 1825

RESULT 5
LOCUS BC011195 2194 bp mRNA linear ROD 30-JUL-2001
DEFINITION Mus musculus, Similar to JM1 protein, clone MGC:18697
IMAGE:4188459, mRNA, complete cds.
ACCESSION BC011195
VERSION BC011195.1 GI:15029928
KEYWORDS house mouse.
SOURCE SCUPCE
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;


```

VERSION      AF378759.1  GI:15987498
KEYWORDS
SOURCE       house mouse.
ORGANISM     Mus musculus

REFERENCE    1 (bases 1 to 5520)
AUTHORS     Carson-Walter,E.B., Vogelstein,B., Kinzler,K.W. and St. Croix,B.
TITLE       Direct Submission
JOURNAL     Submitted (10-MAY-2001) Oncology, Johns Hopkins University, 1650
            Orleans Street, Baltimore, MD 21231, USA

FEATURES
  source     1..5520
             /organism="Mus musculus"
             /db_xref="taxon:10090"
             1..5520
             /gene="Tem5"
             127...204
             /gene="Tem5"
             127...4115
             /gene="Tem5"
             /note="MIEM5: seven-pass transmembrane protein"
             /codon_start=1
             /product="tumor endothelial marker 5 precursor"
             /protein_id="AA111936.1"
             /db_xref="GI:15987499"
             /translation="MPVPPARLLLPCLLLPAGTGPAGCPVPIRGCKSGEPR
             KGLSGAHNARRVCGGGDPEPPDGLPNGITLLSNHKTIGLRNGSFLGLSL
             LEKLDPSNITVOPCAFLGLGKLDLNNRIGCLTSEIFQGLRLLRLNITSGNI
             YSSLOPCVDELPAKLVDFETFLCTDCPLRWLLPNARNHSLQSLERTLCVPSALH
             AHALLSQEQLRCGALHLYLHSLRQVYFGDRLPQCSASYLGNDRIRHWT
             NGAPWESQGLAVLAENLHDCIFITSELTHSHIGWASGENECSYSTVOGNTSKRV
             EIVLETASQCPAERTNNRNGDFRMPRTLAGITAYQSCLOYPTSPVLSGGAGPTPA
             SRRCDRAGRPDGYSHCLYNDITRYLTVFLMPINASHALTLAHLQRYTFAEAF
             SDMMVYVQAAMIOKFLGYVDQIKELVEMVMDASNLMLVDEHLLWLAQREDKACSGI
             VGALERIGGAALSPHQAHOHSVNSRVALEAYLKPHSYVGLTCTAFORREVGVSGAOP
             SSVGODAPVEPEPLADQOLPERCTIGRPNISLSFHKNSVALASIQLPSPSLTILA
             ALAPVPDCTQLQLVFNRLFRSHGNTSRPGAAGPRFGVATPVITAGTSGCGV
             GNLTPEVAVSLRHWAEGADPAAWMDQDGGWSSEGCRLRYSQPNVSSLYCOHLGHW
             AVLMLNAFPREAGSGAGHPVYVYPCITALLLCFTITITILNHSIHVSRLGWHM
             WPAPTLEGEAAPGPRPMLRFLYDIAGGIPLIICGITAAVNHVNRHSPVCMVLRPP
             SLGAFYIPVALILPITWYELCAGLHLSRVAONPKOGRNISLEPGEELRGSTRRSS
             GYLLNDSCGLLATYSAGVGTAPPEDGDGYSPGVOLGALMI THEFLYLNWACGALAY
             SORWLPVRCCLTGVAAALGLFVTHHCARRDRVRASNRACCPASPASASHYPARA
             LPTATEDGSPVLGGEPASLKSQSSGSGRAPPPCKLTNLQVQSVCEASVAARGDG
             EPEPTGSGSLAPRHHNLHGRVRHRSRAKHGAGETGGKSRKLALRAGTSGAPEL
             LSSSEGLHNSPDSYPGSSRNSPDGLPLEGEPMLTPEGSDTSAAPIAETIRPGOR
             RSASRDLKSGSALERESKRRSPYPLNTISLNGAPKGGKYELASVTGAELAGGSMKT
             GLKKSFTIV"
             205...4113
             /gene="Tem5"
             /product="tumor endothelial marker 5"
             2413...2478
             /gene="Tem5"
             /note="Region: predicted transmembrane domain"
             2518...2583
             /gene="Tem5"
             /note="Region: predicted transmembrane domain"
             2599...2664
             /gene="Tem5"
             /note="Region: predicted transmembrane domain"
             2749...2814
             /gene="Tem5"
             /note="Region: predicted transmembrane domain"

```

```

misc_feature 2853..2931
             /gene="Tem5"
             /note="Region: predicted transmembrane domain"
             3151..3207
             /gene="Tem5"
             /note="Region: predicted transmembrane domain"
             3244..3312
             /gene="Tem5"
             /note="Region: predicted transmembrane domain"

BASE COUNT 1125 a 1631 c 1600 g 1164 t

ORIGIN
Query Match 7.3% Score 42; DB 10; Length 5520;
Best Local Similarity 56.5% ; Pred. No. 2;
Matches 78; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Q: 431 qgcagaattcttgagatcagccatttggtacaaaaagatttttaagcttttatgtt 490
  || || || || || || || || || || || || || || || || || || || ||
Db 5383 GGGAAAGGGGAGGAGGAGGATGAAGTCCCAAGAAACACACATTTTAAAGCTCTGTTT 5442
  || || || || || || || || || || || || || || || || || || || ||

Q: 491 ataccatggagccatagaaagctatgattgtttaagaactatttttaagtgtccaga 550
  || || || || || || || || || || || || || || || || || || || ||
Db 5443 ATACAATAGAAATGTTTCCAGCAGATGCTCTCTTTTAAATATATATAATCTTGCAA 5502

Q: 551 cccaaaaaagaaaaaaa 568
  || || || || || || || || || || || || || || || || || || || ||
Db 5503 AAAAAAAGAAAAA 5520

RESULT 14
AC093307/c 153000 bp DNA linear HTG 16-AUG-2001
LOCUS Homo sapiens chromosome 5 clone RP11-810C16, WORKING DRAFT
DEFINITION AC093307
SEQUENCE, 13 unordered pieces.
ACCESSION AC093307
VERSION AC093307.1 GI:15193441
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 153000)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 153000)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1605608
Center clone name: RPC1-11_810C16
-----
Summary Statistics
Consensus quality: 142440 bases at least Q40
Consensus quality: 148623 bases at least Q30
Consensus quality: 149732 bases at least Q20
Estimated insert size: 181920; agarose-fp estimation
Estimated insert size: 151800; sum-of-contigs estimation
Quality coverage: 12.65 in Q20 bases; agarose-fp estimation
Quality coverage: 15.16 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
  * consists of 13 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.

```


Tue Apr 30 14:18:13 2002

	/note="assembly_fragment"	
misc_feature	8471..11234	
	/note="assembly_fragment"	
misc_feature	11335..14011	
	/note="assembly_fragment"	
misc_feature	14112..18113	
	/note="assembly_fragment"	
misc_feature	18214..21222	
	/note="assembly_fragment"	
misc_feature	21323..23848	
	/note="assembly_fragment"	
misc_feature	23949..26966	
	/note="assembly_fragment"	
misc_feature	27067..30515	
	/note="assembly_fragment"	
misc_feature	30616..34597	
	/note="assembly_fragment"	
misc_feature	34698..38352	
	/note="assembly_fragment"	
	clone_end:T7	
	vector_side:left*	
misc_feature	38453..43012	
	/note="assembly_fragment"	
misc_feature	43113..48537	
	/note="assembly_fragment"	
misc_feature	48638..54443	
	/note="assembly_fragment"	
	clone_end:SP6	
	vector_side:left*	
misc_feature	54544..59781	
	/note="assembly_fragment"	
misc_feature	59882..69303	
	/note="assembly_fragment"	
misc_feature	69404..78194	
	/note="assembly_fragment"	
misc_feature	78295..87345	
	/note="assembly_fragment"	
misc_feature	87446..96394	
	/note="assembly_fragment"	
misc_feature	96495..104869	
	/note="assembly_fragment"	
misc_feature	104970..121516	
	/note="assembly_fragment"	
misc_feature	121617..148824	
BASF_COUNT	43002 a 30679 c 30099 g 42736 t 2308 others	
ORI::IN		
Ory Match	7.5%; Score 43; DB 2; Length 148824;	
Best Local Similarity	51.3%; Pred.No. 0.99;	Gaps 0:
Matches 100; Conservative	0; Mismatches 95; Indels 0;	Gaps 0:
Oy 381	tgaagagagatgaaatttcaagtgtgcaggcaggcgctccagtatggcaaagqtt 440	
Dbl 22222	TGTTTGTTGGGGATAAATTAAAGACGTGTGTGTATGAATAAAAACTGAAGTITGTGTT 122163	
Oy 441	cctaagaatcagcatgtgatcacaaaagattttaaacgcttttatgataaccatgga 500	
Dbl 22162	TTTTGCAGGGCGCAATGAGCATCAAACAATTTTGGAAATTTTAATGTTTCTTCAGC 122103	
Oy 501	gccatagaaggctatgtagtttaagaactattttaagtggtcccagacccaagaagg 560	
Dbl 22102	TCCCTGATGGGCCAGGAIGTGGGTTAAAATTGCTTTCCTGTATTAATAAAAAAAAAA 122043	
Oy 561	aaaaaaaaaaaaa 575	
Dbl 22042	AAAAAAAAAAAAAAAAA 122028	
PES'LT 10		
AL5' / 6503		
LOC'SS	AL5065303	linear PRI 07-NOV-2001

DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Human DNA sequence from clone RP11-63B15 on chromosome 1, complete sequence.	AL506503	GI:16904465	HTG.	human.	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
					1 (bases 1 to 143104)					
					Direct Submission					
					Submitted (06-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk clone requests: clonerequest@sanger.ac.uk					
					On Nov 12, 2001 this sequence version replaced gi:16555337.					
					During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.					
					This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Ss., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at					
					http://www.sanger.ac.uk/projects/C-elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at					
					http://www.sanger.ac.uk/HGP/Chrl					
					RP11-63B15 is from the library RCL1-11.1 constructed by the group of Pieter de Jong. For further details see					
					http://www.chori.org/bacpac/home.htm					
					VECTOR: pBACe3.6					
					IMPORTANT: This sequence is not the entire insert of clone RP11-63B15. It may be shorter because we sequence overlapping sections only once, except for a short overlap.					
					The true left end of clone RP11-63B15 is at 1 in this sequence. The true left end of clone RP11-30701 is at 141105 in this sequence.					

```

FEATURES      Location/Qualifiers
source              .143104
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="1"
    /clone="Rpil-63B15"
    /clone_lib="RPCI-11.1"
BASE COUNT      45848 a 25954 c 24891 g 45411 t
ORIGIN
Query Match          7.4%: Score 42.4; DB 9; Length 143104;
Best Local Similarity 49.1%; Pred No. 1.4;
Matches 112; Conservative 0; Mismatches 116; Indels 0; Gaps 0
Qy: 347 aaagaggccagcttaatagcgcgaataacatgacctgaagaagagagatgaattttcaagctg 405
Db 15206 AATAAAGGCACGTAGGTTTCTAATATCAATGGATTAGAAATAAATCTAAAATGTGAACCAG 15265
Qy: 407 tgaagcaggcagagctcccgatggccaaggttcttgagaatcagccatttgtacaaa 466
Db 15265 ACATACAIGGATGATTCATTTTTTAACAAAGGTGTCATCAGAGTACTCATCGGANAACA 15325
Qy: 467 aaagattttlaaagcgttttatgtttataccatgggcccatagaaggctatggatlgitta 526

```



```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Lodes,M.J., Wang,T., Mohamath,R. and Indirias,C.Y.
TITLE Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0177158-A 245 18-OCT-2001:
CORIXA CORPORATION (US)
FEATURES
source 1..473
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 144 a 105 c 84 g 140 t
ORIGIN
Query Match 54.3%; Score 370; DB 6; Length 473;
Best Local Similarity 99.5%; Pred. No. 1.8e-84;
Matches 392; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Oy 1 acttgataaaagagatccatgaatgaattttatactgcattcctttacattagcca 60
|||||
Db 82 ACTTGATATAAAAGGATATCCATAAIGAATATTTATACATCCATTCATTAGCCA 141
|||||

Oy 61 ctaaatcagttattgcttgatgaagaccctttcacagaatcctatgattgcagcattca 120
|||||
Db 142 CTAATACGTTATGCTTGATGAAGACCCTTTCACAGAATCCCTATGGATTGCAGCATTTCA 201
|||||

Oy 121 cttegtacttcataccatgccttaaaagggcgagtttctcaaaagcagaaacatgcc 180
|||||
Db 202 CTGGCTACTTCATACCCATGCCCTTAAAGAGGGCGAGTTTCTCAAAAGCAGAAACATGCC 261
|||||

Oy 181 gccagttctcaagtttctccttaactccatttgatgaaggcgagctggccccaatg 240
|||||
Db 262 GCCAGTTCTCAAGTTTCTCCTCAACTCCATTTGAATGTAAGGCGAGCTGGCCCCCAATG 321
|||||

Oy 241 tggaggagtcggaacatttctgaattcccatcttctgttcggcgctaaatgacagttt 300
|||||
Db 322 TGGGAGTCGGAACATTTTCTGAATTCCTCAATTTCTTGTTTCGGCGCTAANTGACAGTTT 381
|||||

Oy 301 ctgtcattacttagattcccgatcttcccaaaagtgattgatttacaagagccagcta 360
|||||
Db 382 CTGTCAATTACTAGATT-CCGATCTTCCCAAAGGTGTGATTTTACAAAGAGCCAGCTA 440
|||||

Oy 361 atagccagaatcatgacctgaaagagagatga 394
|||||
Db 441 ATAG-CAGAAATCAIGACCCCTGAAGAGAGATGA 473
|||||

RESULT 7
AX321385
LOCUS AX321385 473 bp DNA linear PAT 15-DEC-2001
DEFINITION Sequence 402 from Patent WO0177158.
ACCESSION AX321385
VERSION AX321385.1 GI:17905282
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Lodes,M.J., Wang,T., Mohamath,R. and Indirias,C.Y.
TITLE Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0177158-A 402 18-OCT-2001:
CORIXA CORPORATION (US)
FEATURES
source 1..473
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 144 a 105 c 84 g 140 t
ORIGIN

```

```

Query Match 64.3%; Score 370; DB 6; Length 473;
Best Local Similarity 99.5%; Pred. No. 1.8e-84;
Matches 392; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Oy 1 acttgataaaagagatccatgaatgaattttatactgcattcctttacattagcca 60
|||||
Db 82 ACTTGATATAAAAGGATATCCATAAIGAATATTTATACATCCATTCATTAGCCA 141
|||||

Oy 61 ctaaatcagttattgcttgatgaagaccctttcacagaatcctatgattgcagcattca 120
|||||
Db 142 CTAATACGTTATGCTTGATGAAGACCCTTTCACAGAATCCCTATGGATTGCAGCATTTCA 201
|||||

Oy 121 cttegtacttcataccatgccttaaaagggcgagtttctcaaaagcagaaacatgcc 180
|||||
Db 202 CTGGCTACTTCATACCCATGCCCTTAAAGAGGGCGAGTTTCTCAAAAGCAGAAACATGCC 261
|||||

Oy 181 gccagttctcaagtttctccttaactccatttgatgaaggcgagctggccccaatg 240
|||||
Db 262 GCCAGTTCTCAAGTTTCTCCTCAACTCCATTTGAATGTAAGGCGAGCTGGCCCCCAATG 321
|||||

Oy 241 tggaggagtcggaacatttctgaattcccatcttctgttcggcgctaaatgacagttt 300
|||||
Db 322 TGGGAGTCGGAACATTTTCTGAATTCCTCAATTTCTTGTTTCGGCGCTAANTGACAGTTT 381
|||||

Oy 301 ctgtcattacttagattcccgatcttcccaaaagtgattgatttacaagagccagcta 360
|||||
Db 382 CTGTCAATTACTAGATT-CCGATCTTCCCAAAGGTGTGATTTTACAAAGAGCCAGCTA 440
|||||

Oy 361 atagccagaatcatgacctgaaagagagatga 394
|||||
Db 441 ATAG-CAGAAATCAIGACCCCTGAAGAGAGATGA 473
|||||

RESULT 8
AX35247/c
LOCUS AX35247 294 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 5756 from Patent WO0194629.
ACCESSION AX35247
VERSION AX35247.1 GI:18125966
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Harrigan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets.
JOURNAL Patent: WO 0194629-A 5756 13-DEC-2001:
Avalon Pharmaceuticals (US)
FEATURES
source 1..294
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 85 a 62 c 51 g 96 t
ORIGIN

```

```

Query Match 43.5%; Score 250; DB 6; Length 294;
Best Local Similarity 98.7%; Pred. No. 1e-53;
Matches 294; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

Oy 264 aattcccatcttctgttcgcgcgctaaagacagtttctgtcttacttagattcccgat 323
|||||
Db 294 AATTCCCATTTTCTTGTTCGGCGCTAATGACAGTTTCTGTCTATTAGATT-CCGAT 236
|||||

Oy 324 ctctcccaaaagtgatttacaagagccagctgaatagccagaataatgacacctga 383
|||||
Db 235 CTCTCCCAAAAGGTGTGATTTTACAAAGAGCCAGCTAATAG-CAGAAATCATGACCTGA 177
|||||

```


Oy 541 gtgttcagaccacaaagagaaa 563
|||||
Db 768 GTGTTCAGACCCAAAAA 790

RESULT 4

LOCUS ARI44162 654 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 27 from patent US 6210883.
ACCESSION ARI44162
VERSION ARI44162.1 GI:15105029
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 654)
AUTHORS Reed, S.G. and Wang, T. Tong.
TITLE Compounds and methods for diagnosis of lung cancer
JOURNAL Patent: US 6210883-A 27 03-APR-2001;
FEATURES Location/Qualifiers
source 1..654
/organism="unknown"
BASE COUNT 218 a 130 c 119 g 180 t 7 others
ORIGIN

Query Match 79.3%; Score 455.8; DB 6; Length 654;
Best Local Similarity 94.6%; Pred. No. 1.8e-106;
Matches 543; Conservative 0; Mismatches 24; Indels 7; Gaps 7;

Oy 1 acttgatataaaaggatccataatgaatattttatactgcacatcctttacattagcca 60
|||||
Db 88 ACTTGATATAAAAGGATAICCATATGAATATTTTATCTGCATCCTTTACATTAGCCA 147
Oy 61 ctaatacgttatgtcttgatgaagaccctttcacagaatcctatggattgcagcatttca 120
|||||
Db 148 CTAATACGTTATTGCTTGATGAAGACCTTTCACAGAACTCTATGTGATTGCAGCATTTCA 207
Oy 121 ctggctacttcatacccatgctttaaaggaggcaggtttctcaaaagcagaaacatgcc 180
|||||
Db 208 CTGGCTACTTTCATACCCATGCTTAAAGAGGGCGAGTTCTCAAAAGCAGAAACATGCC 257
Oy 181 gccagttctcaagtttctcctcaactccatttgaagtgaaggcagctggcccccattg 240
|||||
Db 268 GCCAGTCTCAAGTTTCTCTCACTCACTCCATTGTAATGTAAGGCGAGCTGGCCCCCAATG 327
Oy 241 tggggaggtcgaaacattttctgaattcccatcttctgttcgcggcgttaaatgacagttt 300
|||||
Db 328 TGGGGAGGTCGGAACATTTCTGAATTCCTCATTTCTGTTCGGGCTAAATGACAGTTT 387
Oy 301 ctgtcattacttagatcccgatcttcccaagggtgttgatttacaagagggccagcta 360
|||||
Db 388 CTGTCACTACTTAGATT-CCGATCTTTCCCAAAAGGTGTGATTTCACAAAGAGGCCAGCTA 446
Oy 361 ataggcagaatacatgacctgaagagagatgaatttcaagctgtgagccagcagga 420
|||||
Db 447 ATAG-CAGAAATCATGACCTTGAAGAGAGATGAAA-TTCAAGCTGTGAGCCAGGCAGGA 504
Oy 421 gctccagatggcaagggtcttgagaatcagccatttggtacaaaagatttttaag 480
|||||
Db 505 MCT-CAGTATGGCAAGG-TCTTGACAAATCGCCATTTCGTACAAAAA-AATTTTAAAG 561
Oy 481 cttttatgttataccatgagccatgagaaggctatggtattgtttaagaactattttaa 540
|||||
Db 562 CNTTATGTTTATACCATGGAACCATAGAAANGGCAAGGGAATTTGTTAAGANAATTTTAA 521
Oy 541 gtgttcagaccacaaagagaaaaa 574
|||||
Db 622 GTG-TCCAGACCCANANGAAGAAAAA 554

RESULT 5

ARI76349
LOCUS ARI76349 654 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 27 from patent US 6312695.
ACCESSION ARI76349
VERSION ARI76349.1 GI:17918704
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 654)
AUTHORS Reed, S.G. and Wang, T. Tong.
TITLE Compounds and methods for therapy of lung cancer
JOURNAL Patent: US 6312695-A 27 06-NOV-2001;
FEATURES Location/Qualifiers
source 1..654
/organism="unknown"
BASE COUNT 218 a 130 c 119 g 180 t 7 others
ORIGIN

Query Match 79.3%; Score 455.8; DB 6; Length 654;
Best Local Similarity 94.6%; Pred. No. 1.8e-106;
Matches 543; Conservative 0; Mismatches 24; Indels 7; Gaps 7;

Oy 1 acttgatataaaaggatccataatgaatattttatactgcacatcctttacattagcca 60
|||||
Db 88 ACTTGATATAAAAGGATAICCATATGAATATTTTATCTGCATCCTTTACATTAGCCA 147
Oy 61 ctaatacgttatgtcttgatgaagaccctttcacagaatcctatggattgcagcatttca 120
|||||
Db 148 CTAATACGTTATTGCTTGATGAAGACCTTTCACAGAACTCTATGTGATTGCAGCATTTCA 207
Oy 121 ctggctacttcatacccatgctttaaaggaggcaggtttctcaaaagcagaaacatgcc 180
|||||
Db 208 CTGGCTACTTTCATACCCATGCTTAAAGAGGGCGAGTTCTCAAAAGCAGAAACATGCC 267
Oy 181 gccagttctcaagtttctcctcaactccatttgaagtgaaggcagctggcccccattg 240
|||||
Db 268 GCCAGTCTCAAGTTTCTCTCACTCACTCCATTGTAATGTAAGGCGAGCTGGCCCCCAATG 327
Oy 241 tggggaggtcgaaacattttctgaattcccatcttctgttcgcggcgttaaatgacagttt 300
|||||
Db 328 TGGGGAGGTCGGAACATTTCTGAATTCCTCATTTCTGTTCGGGCTAAATGACAGTTT 387
Oy 301 ctgtcattacttagatcccgatcttcccaagggtgttgatttacaagagggccagcta 360
|||||
Db 388 CTGTCACTACTTAGATT-CCGATCTTTCCCAAAAGGTGTGATTTCACAAAGAGGCCAGCTA 446
Oy 361 ataggcagaatacatgacctgaagagagatgaatttcaagctgtgagccagcagga 420
|||||
Db 447 ATAG-CAGAAATCATGACCTTGAAGAGAGATGAAA-TTCAAGCTGTGAGCCAGGCAGGA 504
Oy 421 gctccagatggcaagggtcttgagaatcagccatttggtacaaaagatttttaag 480
|||||
Db 505 MCT-CAGTATGGCAAGG-TCTTGACAAATCGCCATTTCGTACAAAAA-AATTTTAAAG 561
Oy 481 cttttatgttataccatgagccatgagaaggctatggtattgtttaagaactattttaa 540
|||||
Db 562 CNTTATGTTTATACCATGGAACCATAGAAANGGCAAGGGAATTTGTTAAGANAATTTTAA 621
Oy 541 gtgttcagaccacaaagagaaaaa 574
|||||
Db 622 GTG-TCCAGACCCANANGAAGAAAAA 654

RESULT 6
LOCUS AX321228 473 bp DNA linear PAT 15-DEC-2001
DEFINITION Sequence 245 from Patent WO0177168.
ACCESSION AX321228
VERSION AX321228.1 GI:17904842
KEYWORDS human.

BASE COUNT 51905 a 36985 c 38442 g 56185 t 100 others
ORIGIN

Query Match 91.4%; Score 525.4; DB 2: Length 183617;
Best Local Similarity 99.1%; Pred. No. 2.2e-124;
Matches 570; Conservative 0; Mismatches 1; Indels 4; Gaps 4;

Oy 1 acttgatataaaagagatccataatgaattttatatactgcatctttacattagcca 60
|||||
Db 32567 ACTTGATATAAAAGGATATCCATAATGAATATTTATACGATCCTTTACATAGCCA 32624
Oy 61 ctataatcgttatgttgatgaagacccttccagaatctctatgatgcagcattca 120
|||||
Db 32627 CTAATAACGTTATGCTTGATGAAGACCTTTCACAGAACTCTATGTAATTCAGCATTTCA 32686
Oy 121 ctgtgctactctataccatgccttaaaaggggcagtttctcaaaagcagaaacatgcc 180
|||||
Db 32687 CTGGCTACTTCATACCATGCGCTTAAGAGGGGCGAGTTTTCACAAAGCAGAAACATGCC 32745
Oy 181 gccagttctcaagtttctctcttaactccatttgaaatgaagggcgctggcccccaatg 240
|||||
Db 32747 GCCAGTTCTCAAGTTTCTCTCTTAACCTCAATTTGAATGTAAGGGCAGCTGGCCCCCAATG 32806
Oy 241 tggggaggtccgaacatttctgaattcccatcttctgttcgcgctaaatgacagttt 300
|||||
Db 32807 TGGGAGGTCGACACATTTTCTGAATCCCATTTTCTGTTCGCGCTAAATGACAGATT 32866
Oy 301 ctgtcattactagattcccgatctttcccaaaaggttgatttacaagagggccagcta 360
|||||
Db 32867 CTGTCAATTACTTAGAAT-CCGATCTTTCACAAAGGTTGATTACAAAGAGGCCAGCTA 32925
Oy 361 atagcagaatcatgaacctgaagagagataaatttcaagctotagccagcagga 420
|||||
Db 32926 ATAG-CAGAAATCAAGCTGAAAGAGAGATGAAA-TTCAAGCTGTGAGCCAGGACGGA 32983
Oy 421 gctccagtggaagaggttcttgagaatcagccatttggtacaaaagagatttttaaag 480
|||||
Db 32984 GCT-CAGTATGGCAAGGTTCTTGAGAAATCAGCCATTGTTGTACAAAAGATTTTAAAG 33042
Oy 481 cttttatgttatccatggagcattagaagagctatggattgtttaagaactttttaa 540
|||||
Db 33043 CTTTATGTTATACCATGGAGCCATAGAAGGCTATGGATTGTTTAAAGAACTATTTTAA 33102
Oy 541 gtgtccagagcccaaaagaaaaaataaaaaa 575
|||||
Db 33103 GTGTTCAGACCCCAAAAGCAAAATATAAAAAA 33137

RESULT 3
HUMY291G09
LOCUS HUMY291G09 790 bp mRNA linear PRI 29-AUG-1998
DEFINITION Homo sapiens full length insert cDNA clone y291G09.
ACCESSION AF086101
VERSION AF086101.1 GI:3483446
KEYWORDS FLI_CDNA.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 790)
REFERENCE
AUTHORS Woessner,J., Tan,F., Marra,M., Kucaba,T., Yandell,M., Martin,J.,
Marth,G., Bowles,L., Wylie,T., Boxers,Y., Sleptoe,M., Theising,B.,
Geisel,S., Allen,M., Underwood,K., Chappell,J., Person,B.,
Gibbons,K., Harvey,N., Pape,D., Chamberlain,A., Morales,R.,
Schurk,R., Pitter,E., Kohn,S., Swaller,I., Behymer,K., Hillis,D.,
Wilson,R. and Waterston,R.
Full Clone Sequencing of the Longest Available Member from Each
Unigene Cluster

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 790)
AUTHORS Waterston,R.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (24-AUG-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
SUBMITTED BY:
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:est@watson.wustl.edu

NOTICE: This sequence represents the full insert of this cDNA. No
attempt has been made to verify whether this corresponds to the
full-length of the original mRNA from which it was derived. We
have tried to obtain double-stranded, or double chemistry sequence
across the entire clone, but potentially, there are areas in the
sequence where this level of coverage was not achieved.
Nevertheless, we are confident of the accuracy of this sequence as
all regions of low quality, as defined by PHRAP (P. Green, in
preparation), were visually inspected and edited accordingly. The
consensus quality values for this sequence have been submitted
separately.

The location of this clone is unknown.

FEATURES
Source

Location/Qualifiers
1..790
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:290464"
/clone_lib="Soares_multiple_sclerosis_2NDHMS"
BASE COUNT 255 a 152 c 145 g 238 t
ORIGIN

Query Match 89.0%; Score 511.8; DB 9: Length 790;
Best Local Similarity 98.9%; Pred. No. 8e-121;
Matches 557; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

Oy 1 acttgatataaaagagatccataatgaattttatatactgcatctttacattagcca 60
|||||
Db 232 ACTTGATATAAAAGGATATCCATAATGAATATTTATACGATCCTTTACATAGCCA 291
Oy 61 ctataatcgttatgttgatgaagacccttccagaatctctatgatgcagcattca 120
|||||
Db 292 CTAATAACGTTATGCTTGATGAAGACCTTTCACAGAACTCTATGATTCGACGATTTCA 351
Oy 121 ctgtgctactctataccatgccttaaaaggggcagtttctcaaaagcagaaacatgcc 180
|||||
Db 352 CTGGCTACTTCATACCCATGCTTAAAGAGGGCAGTTTCTCAAAAGCAGAAACATGCC 411
Oy 181 gccagttctcaagtttctctcttaactccatttgaaatgaagggcgctggcccccaatg 240
|||||
Db 412 GCCAGTTCTCAAGTTTCTCTCTTAACCTCAATTTGAATGTAAGGGCAGCTGGCCCCCAATG 471
Oy 241 tggggaggtccgaacatttctgaattcccatcttctgttcgcggtctaaatgacagttt 300
|||||
Db 472 TGGGAGGTCGCAACATTTTCTGAATCCCATTTTCTGTTCGCGCTAAATGACAGATT 531
Oy 301 ctgtcattactagattcccgatctttcccaaaaggttgatttacaagagggccagcta 360
|||||
Db 532 CTGTCAATTACTTAGATT-CCGATCTTTCACAAAGGTTGATTTACAAAGAGGCCAGCTA 590
Oy 361 atagcagaatcatgaacctgaagagagataaatttcaagctotagccagcagga 420
|||||
Db 591 ATAG-CAGAAATCAATCACCCTTGAAGAGAGATGAAA-TTCAAGCTGTGAGCCAGGACGA 648
Oy 421 gctccagtggaagaggttcttgagaatcagccatttggtacaaaagagatttttaaag 480
|||||
Db 649 GCT-CAGTATGGCAAGGTTCTTGAGAAATCAGCCATTGTTGTACAAAAGATTTTAAAG 707
Oy 481 cttttatgttatccatggagcattagaagagctatggattgtttaagaactttttaa 540
|||||
Db 708 CTTTATGTTATACCATGGAGCCATAGAAGGCTATGGATTGTTTAAAGAACTATTTTAA 767

repeat_region /note="L2 repeat: matches 2580. 2248 of consensus" 42426. 42626
repeat_region /note="L2 repeat: matches 2023. 2228 of consensus" 43755. 43899
repeat_region /note="MULTI1 repeat: matches 59. 205 of consensus" 43997. 45345
repeat_region /note="L1P8 repeat: matches 4813. 5161 of consensus" 45917. 46305
misc_feature /note="match: GSS: Em:A0075236" 46515. 46785
repeat_region /note="L2 repeat: matches 2408. 2468 of consensus" 47901. 48024
repeat_region /note="AluJb repeat: matches 3. 125 of consensus" 49747. 49910
repeat_region /note="MIR repeat: matches 56. 244 of consensus" 50423. 50675
repeat_region /note="L1MB3 repeat: matches 5926. 5176 of consensus" 51366. 51550
repeat_region /note="AluJ repeat: matches 127. 131 of consensus" 53273. 53576
repeat_region /note="AluSx repeat: matches 3. 219 of consensus" 54377. 54432
repeat_region /note="MIR repeat: matches 53. 112 of consensus" 54496. 54612
repeat_region /note="MIR repeat: matches 61. 184 of consensus" 54927. 55102
repeat_region /note="MIR repeat: matches 3. 183 of consensus" 55627. 55690
misc_feature /note="2 copies 32 mer 92% conserved" complement(55692..55879)
misc_feature /note="match: STS: Em:L30443" complement(55692..55873)
misc_feature /note="match: STS: Em:L18003" 91.48; Score 525.4; DB 9; Length 182914; Best Local Similarity 95.18; Pred. No. 2.2e-124; Matches 570; Conservative 0; Mismatches 1; Indels 4; Gaps 4;

Oy 1 acttgatataaaagatatacctaataatgatttttttatactgcatcttttatacattagcca 50
Db 173021 ACTTGATATAAAAGATATACCAATCAATCAATATTTTATCTGCATCTTTACATTAGCCA 173080

Oy 61 ctaataacgtattcttgatgaagacaccttcacagaactctatgatttcagcatttca 120
Db 173081 CTAATAACGTTATTCCTTGATGAAGACCTTTCACAGAATCTCTATGTAITGCAGCATTTCA 173140

Oy 121 ctggctacttcataccatgcctctaaagaggcgagttcttcaaaagcagaacaatgcc 180
Db 173141 CTGGCTACTTTCATACCAATGCTTAAAGAGGGCGAGTTTCTCAAAGCAGAAACATGCC 173200

Oy 181 gccagttctcaagtttctctactcctcatttgaaatgaaggcagctgagcccccattg 240
Db 173201 GCCAGTTCTCAAGTTTCTCTCACTCTCATTTGAATGAAGGGGAGCTGTGGCCCCCAATG 173260

Oy 241 tggggagggtccgaacatttcttgaattccccatttcttctgttcgcgcctaaatgacagttt 300
Db 173261 TGGGGAGGTCCGACAAATTTCTGAATCCCAATTTCTTGTTCGGCGCTAAATGACAGTTT 173320

Oy 301 ctctcattacttagattcccgatcttcccaagagtgattgatttcaagagggccaacta 360
Db 173321 CTCTCATTTACTTAGATT-CCGATCTTCTCCAAAGGTGTTGATTATTAAGAGGGCCAGCTA 173379

Oy 361 atagccagaaatcatgacctgaagagagatgaatttcaagcctatgagccaggcaaga 420
Db 173380 ATAG-CAGAAATCATGACCTTGAAGAGATGANA-TTCAGGCTGTAGCCAGGCACAGA 173437

Oy 421 gctccagatggcaaggttcttgagaatcagccatttggttcacaaagaagattttanaag 480
Db 173438 GCT-CAGTATGGCAAGGTTCTTGAGAAATCAGCCATTTGGTACAAATAAGATTTTAAAG 173495

Oy 481 cttttatgttataccatgagccatagaagagctatggattgttttgaacactatttttaa 540
Db 173497 CTTTATGTATTATACCATGGAGCCATAGAAAGGCTATGGATTGTTTAAAGACTATTTTANA 173556

Oy 541 gtattccagaccacaaagagaaaaa 575
Db 173557 GTGTTCCAGACCCAAAAGAAAATAAATAAANA 173591

RESULT 2
AC007524 183617 bp DNA linear HTG 14-OCT-2001
LOCUS Homo sapiens chromosome 9 clone RP11-30A1 map 9, *** SEQUENCING IN
DEFINITION PROGRESS ***, 2 ordered pieces.
ACCESSION AC007524 G1:16118190
VERSION HTG: HTGS-PHASE2; HTGS-FULLTOP; HTGS-ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 183617)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 9, clone RP11-30A1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 183617)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dextar,K., Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Pollara,V., Riley,A., Roberts,D., Roy,A., Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A., Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo.A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (11-MAY-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 14, 2001 this sequence version replaced gi:14579797.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L602
Center clone name: 30_A_1

NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.
1 47086: contig of 47086 bp in length
47087 47186: gap of 100 bp
47187 183617: contig of 136431 bp in length.
----- Location/Qualifiers
1. 183617
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="9"
/clone="RP11-30A1"

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr9>
RP11-301F14 is from the library RP11-11.2 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>
VFC10R: PBA03.5

This sequence is the entire insert of clone RP11-301F14. The true right end of clone RP11-301M4 is at 35779 in this sequence.

FEATURES

source

Location/Qualifiers
1..182914
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-301F14"
/clone_lib="RP11-11.2"
1045..1250

/note="MER58A repeat: matches 2..224 of consensus"
1397..1432

/note="18 copies 2 mer ac 97% conserved"
1399..1434

/note="9 copies 4 mer acac 97% conserved"
2899..3234

/note="HE1B repeat: matches 2..364 of consensus"
3612..3595

/note="MLT11 repeat: matches 316..410 of consensus"
4054..4557

/note="L1MD2 repeat: matches 5855..6341 of consensus"
4559..5116

/note="MLR2D repeat: matches 3..553 of consensus"
5119..5782

/note="HERVL repeat: matches 5097..5757 of consensus"
5794..5852

/note="L1PA5 repeat: matches 5070..6141 of consensus"
5853..7814

/note="HERVL repeat: matches 4126..5111 of consensus"
7847..9162

/note="HERVL repeat: matches 2867..4181 of consensus"
9163..9573

/note="L1MA8 repeat: matches 5773..6289 of consensus"
9574..9584

/note="HERVL repeat: matches 2857..2867 of consensus"
9587..10596

/note="L1MA5A repeat: matches 5470..6293 of consensus"
10591..11052

/note="L1 repeat: matches 4509..4829 of consensus"
10953..11266

/note="AluX repeat: matches 1..312 of consensus"
11267..11381

/note="L1 repeat: matches 4387..4509 of consensus"
11405..12429

/note="L1MD2 repeat: matches 4222..5939 of consensus"
14192..14223

/note="8 copies 4 mer ctct 87% conserved"
14733..15026

/note="AluX repeat: matches 1..296 of consensus"
15827..16136

/note="Aluub repeat: matches 5..309 of consensus"

repeat_region

repeat_region

misc_feature

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

15653..16989

/note="L2 repeat: matches 2330..2670 of consensus"

17784..18071

/note="AluSq repeat: matches 25..311 of consensus"

complement(18427..18978)

/note="match: GSS: Em:AQ683961"

19374..19656

/note="AluSc repeat: matches 1..286 of consensus"

20117..20584

/note="MER56B repeat: matches 1..476 of consensus"

20860..21033

/note="MLT1J repeat: matches 90..262 of consensus"

22186..22489

/note="AluSq repeat: matches 1..307 of consensus"

22552..23422

/note="Tigger3b repeat: matches 365..1221 of consensus"

23423..23725

/note="AluY repeat: matches 11..311 of consensus"

23726..24055

/note="Tigger3b repeat: matches 1..365 of consensus"

complement(23903..24656)

/note="match: GSS: Em:AQ746687"

24480..25040

/note="L1ME1 repeat: matches 5460..6012 of consensus"

25041..25299

/note="AluY repeat: matches 54..309 of consensus"

25300..25652

/note="L1ME1 repeat: matches 5107..5450 of consensus"

25650..26011

/note="L1M4 repeat: matches 4060..4424 of consensus"

26012..26094

/note="L1MA3 repeat: matches 5928..6017 of consensus"

26099..27026

/note="L1M4 repeat: matches 2892..3829 of consensus"

26989..27238

/note="L1MEC repeat: matches 477..739 of consensus"

28100..28295

/note="L2 repeat: matches 2540..2750 of consensus"

29076..31601

/note="Cpg Island"

/evidence=not_experimental

30253..30280

/note="14 copies 2 mer ca 92% conserved"

30446..30483

/note="19 copies 2 mer tg 78% conserved"

complement(32749..33229)

/note="match: STS: Em:G57804"

match: GSS: Em:AQ343604"

33339..33866

/note="match: GSS: Em:AQ546349"

34099..34545

/note="match: GSS: Em:AQ706050"

34160..34506

/note="match: GSS: Em:AQ630236"

34917..35007

/note="L1MB2 repeat: matches 6069..6166 of consensus"

36103..36162

/note="30 copies 2 mer tt 81% conserved"

36117..36161

/note="5 copies 9 mer tctttttt 86% conserved"

36886..37008

/note="M1R repeat: matches 33..145 of consensus"

37551..37605

/note="MADE1 repeat: matches 1..80 of consensus"

37754..37795

/note="8 copies 4 mer tgtg 93% conserved"

39699..40035

/note="MLT1A1 repeat: matches 1..362 of consensus"

complement(40477..40847)

/note="match: GSS: Em:AQ024170"

41352..42246

/note="L2 repeat: matches 1042..2013 of consensus"

42258..42437

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM - nucleic - nucleic search, using sw model

Run on: April 29, 2002, 20:03:59 ; Search time 8020.64 Seconds
(without alignments)
1500.224 Million cell updates/sec

Title: US-09-248-178-65
Perfect score: 575
Sequence: 1 actgataataaaaggatat.....aaaggaataaaaaaa 575

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Sea-ched: * 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 100%
Listing first 45 summaries

Post-processing: Minimum Match 0%

Dat base : GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
.....					

1	525.4	91.4	182914	9	AL390777	Human DNA
2	525.4	91.4	183617	2	AC007524	Homo sapi
3	511.8	89.0	790	9	HUMY291G09	Homo sapi
4	455.8	79.3	654	6	AR144162	Sequence
5	455.8	79.3	654	6	AR176349	Sequence
6	370	64.3	473	6	AX321228	Sequence
7	370	64.3	473	6	AX321365	Sequence
8	250	43.5	294	6	AX335247	Sequence
9	43	7.5	148824	2	AC025640	Homo sapi
10	42.4	7.4	143104	9	AL606503	Human DNA
11	42.4	7.4	151186	2	AC068750	Homo sapi
12	42.2	7.3	110146	2	AC022421	Homo sapi
13	42	7.3	5520	10	AF378759	Mus muscu
14	42	7.3	153000	2	AC093307	Homo sapi
15	42	7.3	173294	2	AC109447	Homo sapi
16	41.4	7.2	160666	2	AC105755	Homo sapi
17	41.4	7.2	168447	2	AL671985	Homo sapi
18	41.4	7.2	171962	2	AC013537	Homo sapi
19	41.4	7.2	175488	2	AC013556	Homo sapi
20	41.2	7.2	199551	2	AC006281	Plasmodiu
21	41.2	7.2	204182	2	AL671870	Mus muscu
22	41	7.1	2515	10	BC006577	Mus muscu
23	40.6	7.1	281497	2	AC108494	Homo sapi
24	40.6	7.1	311018	2	AC093684	Homo sapi
25	39.2	6.8	182150	9	HSG278N14	Human DNA
26	38.8	6.7	2379	3	DD31	D. discoideu
27	38.8	6.7	30614	9	AB045362	Homo sapi
28	38.8	6.7	146423	2	AC073856	Homo sapi
29	38.8	6.7	180742	9	AL359753	Human DNA
30	38.8	6.7	194529	2	AC036146	Mus muscu
31	38.8	6.7	300000	9	AP002534	Homo sapi
32	38.6	6.7	91825	2	AC019854	Drosophil
33	38.6	6.7	200978	3	AC010060	Drosophil
34	38.6	6.7	274585	3	AE003533	Drosophil
35	38.4	6.7	1134	9	HSM802368	Homo sapi
36	38.4	6.7	105100	9	HS67A5	Human DNA
37	38.4	6.7	125020	9	AF429315	Homo sapi
38	38.4	6.7	293181	1	AP001119	Buchnera
39	38.2	6.6	28586	3	CEC1A6	Caenorhabdi
40	38.2	6.6	34584	8	SPCC1795	S.pombe c
41	38.2	6.6	171478	9	AC016542	Homo sapi
42	38.2	6.6	219728	2	AL645987	Mus muscu
43	38	6.6	108908	3	PFMAL3P8	Plasmodiu
44	38	6.6	130764	2	AC094324	Rattus no
45	37.8	6.6	5173	9	AF426432	Homo sapi

ALIGNMENTS

RESULT 1

AL390777

LOCUS

DEFINITION

AL390777

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL390777 Human DNA sequence from clone RP11-301f14 on chromosome 9, complete sequence.

AL390777

AL390777.13 GI:14329989

HTG.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 182914)

Brown, J.

Direct Submission

Submitted (02-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Jun 8, 2001 this sequence version replaced qi:14727213.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

Tue Apr 30 14:17:53 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OH nucleic - nucleic search, using sw model

Run on: April 29, 2002, 19:55:33 : Search time 202.96 seconds
(without alignments)
462.318 Million cell updates/sec

Title: US-09-248-178-59
Perfect score: 382
Sequence: 1 aggcggagcagagctaaa.....attaaaaaa 382

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues 767066

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents.NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

F-Result No.	Score	Query Match	Length	ID	Description
1	367.4	96.2	384	US-09-030-607-183	Sequence 183, App
2	367.4	96.2	384	US-09-439-313-183	Sequence 82, Appl
3	354.8	92.9	383	US-09-030-607-82	Sequence 82, Appl
4	354.8	92.9	383	US-09-439-313-82	Sequence 73, Appl
5	354.8	92.9	499	US-09-020-956-73	Sequence 73, Appl
6	335.6	87.9	499	US-09-030-607-73	Sequence 1, Appl
7	335.6	87.9	499	US-09-439-313-73	Sequence 1, Appl
8	335.6	87.9	499	US-08-718-661-1	Sequence 1, Appl
9	45.4	11.9	2790	US-08-317-522A-1	Sequence 1, Appl
10	43.6	11.4	2524	US-08-439-818A-1	Sequence 1, Appl
11	43.6	11.4	2524	US-08-751-965-1	Sequence 1, Appl
12	43.6	11.4	2524	US-08-738-975-1	Sequence 1, Appl
13	43.6	11.4	2524	US-08-728-626-1	Sequence 1, Appl
14	43.6	11.4	2524	US-08-808-599A-1	Sequence 1, Appl
15	43.6	11.4	2524	US-09-128-155-17	Sequence 1, Appl
16	43.2	11.3	176373	US-08-317-522A-1	Sequence 1, Appl
17	42.6	11.2	2524	US-08-439-818A-1	Sequence 1, Appl
18	42.6	11.2	2524	US-08-751-965-1	Sequence 1, Appl
19	42.6	11.2	2524	US-08-738-975-1	Sequence 1, Appl
20	42.6	11.2	2524	US-08-728-626-1	Sequence 1, Appl
21	42.6	11.2	2524	US-08-808-599A-1	Sequence 1, Appl
22	42.6	11.2	2524	US-08-128-155-16	Sequence 1, Appl
23	42.6	11.0	3483	US-08-08-599A-23	Sequence 1, Appl
24	41.4	10.8	152331	US-08-128-155-16	Sequence 1, Appl
25	40	10.5	5159	US-08-458-240-3	Sequence 1, Appl
26	40	10.5	5159	US-08-458-240-3	Sequence 1, Appl
27	40	10.5	5159	PCT-US93-03993-3	Sequence 1, Appl

Sequence 3, Appl
Sequence 21, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 2, Appl
Sequence 1, Appl
Sequence 2, Appl
Sequence 4, Appl
Sequence 198, Appl
Sequence 1, Appl
Sequence 45, Appl
Sequence 45, Appl
Sequence 34, Appl
Sequence 51, Appl
Sequence 86, Appl
Sequence 59, Appl
Sequence 86, Appl

ALIGNMENTS

RESULT 1
US-09-030-607-183

Sequence 183, Application US/09030607
Patent No. 6262245
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
NUMBER OF SEQUENCES: 224
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: WA USA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/030.607
FILING DATE: 25-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.427C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 522-4900
TELEFAX: (206) 582-6031
INFORMATION FOR SEQ ID NO: 183:
SEQUENCE CHARACTERISTICS:
LENGTH: 384 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-030-607-183

Query Match 96.2% Score 367.4; DB 4; Length 384;
Best Local Similarity 99.0%; Pred. No. 5.1e-99; Indels 1; Gaps 1;
Matches 379; Conservative 1; Mismatches 2;

Q: 1 aggcggagcagagctaaaagccaaagccagag-aaagtgcagtgccagcactgagcc 59
|||||
1 AGCGGGGAGCAGAGCTAAAGCCAAAGCCAAAGAGAGTGGCTGCGCAGCTGGTGGCC 60

DB 1

Q: 50 agtaccagttaccataaacagtgccagtgccagcaccagtggtggttccagtgct 119

RESULT

Quarry Match

Best: Local Similarity 96.28; Score 367.4; DB 4; Length 384;
Matches 379; Conservative 1.0; Pred. No. 5.1e-99;

	Conservative	1: Mismatches	2: Indels	1: Gaps
OY	1	aggcggagcagaagctaaagccaaagcccaag-agaagtgcagctccagcactggtgcc	1	
Db	1	aggcggagcagaagctaaagccaaagcccaagagaagtggcagctgggtgcc	59	
OY	60	agtaccagctaccataaacagctgcagctgcagctccagaccagctggctcagctt	50	
Db	61	agtaccagctaccataaacagctgcagctgcagctgcagaccagctggctcagctt	113	
OY	120	ggctccagcctgaccgcactctcacattggcctcttcgtgccttgggtggagctgct	120	
Db	21	ggctccagcctgaccgcactctcacattggcctcttcgtgccttgggtggagctgct	179	
OY	20	gccagccagctggcagctctagtgccctggcttctctctacagctagatttagatat	180	
Db	31	gccagccagctggcagctctagtgccctggcttctctctacagctagatttagatat	239	
..		ggcagccagctggcagctctagtgccctggcttctctctacagctagatttagatat	240	

Query Match	92.9%	Score 354.8	DB 4	Length 383
Best Local Similarity	98.7%	pred. No. 2.6e-95		
Matches 378	Conservative	0	Mismatches	3
			Indels	2
OY	1	agcgggagcagaagctaaagccaaagcccaag-aagatgacgagtcgcagcactgacc	59	Gaps
Db	1	AGGCGCGAGCAGAGCTAAAGCCAAAGCCCAAGAAAGATGGCAGCTGGTGCC	60	
OY	50	agtcaccatgaccataaacagtgccagtgccagtcgcagcaccagtgagtgcttcagtcct	119	
Db	61	AGTACCGATACCAATAACA-TGCCAGTGGCAGTGCAGCACCAGTGGTGGCTTCAGTGCT	119	
OY	120	agtgccagcctgaccgcacactctcacattggactcttcgctggccttgatggagactcgtt	179	
Db	120	GGTGCCAGCCTGACGCCACTCTCAGATTGGGCTTTCGCTGGCCTTGGTGGAGCTGGT	179	

Accession	Sequence	Length
Db	180 gccagccagtgccagctctggctctggctctctacaaggagatcttcaatat	239
Oy	240 tgttaacctgcagctcttctctcaagccagggtgcatctctcaaaacactactcaaca	299
Db	240 tgttaacctgccagctcttctctcaagccagggtgcatctctcaaaacactactcaaca	299
Oy	300 cagcactctaggcagccactatcaatcaattggaagttgacactctcacattaactctatt	359
Db	300 cagcactctngcagccactatcaatcaattggaagttgacactctgcattcaactctatt	359
Oy	360 gccattaaaaaaaaaaaaaaaaa	382
Db	360 gccatttcaaaaaaaaaaaaaaa	382

DEPT. T 6

US-02-020-956-73
: Sequence 73, Application US/09020956
: Patent No. 6261562
: GENERAL INFORMATION:
: APPLICANT: Xu, JiangChun
: APPLICANT: Dillin, Davin C.
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FOR
: NUMBER OF SEQUENCES: 178
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED and BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/020.956
: FILING DATE: 09-FEB-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Makl, David J.
: REGISTRATION NUMBER: 31,392
: REFERENCE/DOCKET NUMBER: 210121.427C2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 522-4900
: TELEFAX: (206) 582-6031
: INFORMATION FOR SEQ ID NO: 73:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 499 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
US-02-020-956-73

	Query Match	87.9%	Score 335.5	DB 4	Length 499
	Best Local Similarity	98.8%	Pred. No. 1.3e-89		
	Matches 318	Conservative	0	Mismatches 4	Indels 0
y	41	cagtgccagcactggtgccagtcaccagtcaccataacagtgccagtcaccatgccagcac	100		
b	1	CAGTGCCAGCAGCTGCTGCCAGTACCAGTACCAATACCAATGCCAGTCCAGTGGCAGCAC	50		
y	101	cagtggtgtgcctcagtcgtgtaccagagcctgacccgacactctcacatttgagctcttcgc	150		
b	51	CAGTGGTGGCTTCAGTCTGGTSCCAGGCTGACGGCCACTCTCACATTTGGGCTCTCGC	120		
	161	tagcctcttgatgaactgattaccagacaccagtcacagctctgctgctctgattctctccta	220		

[illegible]

RESULT 7

US-09-030-607-73
: Sequence 73, Application US/09030607
: Patent No. 5262245
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: APPLICANT: Dillon, Davin C.
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
: NUMBER OF SEQUENCES: 224
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED and BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/030.607
: FILING DATE: 25-FEB-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Makl, David J.
: REGISTRATION NUMBER: 31,392
: REFERENCE/DOCKET NUMBER: 210121.427C3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO: 73:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 499 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
US-09-030-607-73

[illegible]

[illegible]

RESULT a

```

US-0439-313-73
: sequence 73, Application US/09439313
: Patent No. 5329505
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: APPLICANT: Dillon, David C.
: APPLICANT: Mitcham, Jennifer L.
: APPLICANT: Harlocker, Susan Louise
: APPLICANT: Jiang Yuqi
: APPLICANT: Reed, Steven G.
: APPLICANT: Kalos, Michael
: APPLICANT: Fanger, Gary
: APPLICANT: Retter, Mark
: APPLICANT: Solk, John
: APPLICANT: Day, Craig
: TITLE OF INVENTION: COMPOSITIONS AND
: TITLE OF INVENTION: DIAGNOSIS OF P
: FILE REFERENCE: 210121.427C9
: CURRENT APPLICATION NUMBER: US/09/4/4
: CURRENT FILING DATE: 1999-11-12
: NUMBER OF SEQ ID NOS: 575
: SOFTWARE: FastSeq for Windows Versi
: SEQ ID NO 73
: LENGTH: 499
: TYPE: DNA
: ORGANISM: Homo sapien
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)..(499)
: OTHER INFORMATION: n = A,T,C or G
US-0439-313-73

```

	Query Match	87.94:	Score 335.6:	DB 4:	Length 499:
	Post Local Similarity	98.8%:	Pred. No. 1.3e-89:		
Matches 338:	Conservative	0:	Mismatches	Indels	Gaps
Oy	41	cagtgccagcactggtgcagttaccagtaccataaacagtcgcaatgccagtcggcgcttgcac	100		
Db	1	cagtgccagcactggtgcagttaccagtaccataaacagtcgcaatgccagtcggcgcttgcac	60		
Oy	101	cagtggtggcttcadgtgtgtgcagcctgaaccgcaccttcacatttgggtctttcgc	160		
Db	61	cagtggtgcttcagtggtgtgcagcctgaccgcaccttcacatttgggtctttcgc	120		
Oy	161	tggccttgttgagctgtgtgcagcacccagtgcgagctctgatgctgtgggtttctcta	220		
Db	121	tggccttgttgagctgtgtgcagcaccagtggcagctctggtgctgtgtttctcta	180		
Oy	221	caagtgaattttaatatattgttaactctgcagctctttctcttcaagccagggtgcac	280		
Db	181	caagtgaattttaatatattgttaactctgcagctctttctcttcaagccagggtgcac	240		
Oy	281	ctcagaacctactcaacacagcactctaggcagccactatcaataattgaattgaca	340		
Db	241	ctcagaacctactcaacacagcactctaggcagccactatcaataattgaattgaca	300		

```

QY: 3:1 ctcgcattaaatctatttgcattaaataaaaaa 382
      ||||| ||||| ||||| ||||| ||||| |||||
Db 3:1 ctcgcattaaatctatttgcatttgcataaaaaa 342

RESULT 9
US-08-718-661-1
: Sequence 1, Application US/08718661
: Patent No. 5876972
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: Nucleic acid molecules coding
: for the production of tumor suppressor proteins and
: methods of use
: NUMBER OF SEQUENCES: 15
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30 (
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/718.661
: FILING DATE:
: CLASSIFICATION: 530
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2790 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 542..2545
: US-08-718-661-1

```

Query Ma

```

Query Match      11.9%: Score 45.4: DB 2: Length 2790:
Best Local Similarity 61.3%: Pred. No. 0.00071:
Matches 71: Conservative 0: Mismatches 46: Indels 0: Gaps 0:

y 3 gcggagcagaagctaaagcccaagagatggcagtcgccagcactggtgccaat 62
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 2275 GCAGCAGCCTGTATCCTGTAGCCAGAGCAGAGCAGCAGCAGCAGCAGCAG 2334
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

y 53 accagtcaccaataacagtcgcagtcgcagtcgccagtcggcttcagtcgtgg 121
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 2335 cccggcagccggacccgacgacgacgacgacgacgacgacgacgacgacgacg 2393
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT

```

: RESOLVED TO :
: US-08-317-522A-1 :
: Sequence 1, Application US/08317522A :
: Patent No. 5599918 :
: GENERAL INFORMATION :
: APPLICANT: Fukuda, Michiko N. :
: TITLE OF INVENTION: Trophinin and Trophinin-Assisting :
: Proteins :
: TITLE OF INVENTION: Proteins :
: NUMBER OF SEQUENCES: 13 :
: COPIES OF SEQUENCE ADDRESS: :
: ADDRESSEE: Campbell and Flores :
: STREET: 4370 La Jolla Village Drive, Suite 700 :
: CITY: San Diego :
: STATE: California :
: COUNTRY: USA :
: ZIP: 92122 :
: COMPUTER READABLE FORM: :
: MEDIUM TYPE: Floppy disk :
: COMPUTER: IBM PC compatible :
: OPERATING SYSTEM: PC-DOS/MS-DOS :
: SOFTWARE: Patent In Release #1.0, Version #1.25 :
: CURRENT APPLICATION DATA: :
: APPLICATION NUMBER: US/08/317.522A :

```


Job time: 22736 sec

```

Db 931 AACACGACTGGTTTGGTGGTGTATGAGCA 964
      | | | | | | | | | | | | | | | | | |
RESULT 15
US-08-808-599A-1
: Sequence 1, Application US/08808599A
: Patent No. 6111089
: GENERAL INFORMATION:
: APPLICANT: Fukuda, Michiko N.
: TITLE OF INVENTION: Trophinin, Trophinin-Assisting
: TITLE OF INVENTION: Proteins and Methods to Inhibit Implantation
: NUMBER OF SEQUENCES: 41
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Campbell & Flores LLP
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/808,599A
: FILING DATE: 28-FEB-1997
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/317,522
: FILING DATE: 04-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/439,818
: FILING DATE: 12-MAY-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LA 2256
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2524 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 28..2275
: US-08-808-599A-1

Query Match 11.4%; Score 43.6; DB 3; Length 2524;
Best Local Similarity 55.2%; Pred. No. 0.0023;
Matches 85; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 36 agtggcagtgccagcactggctggtgcccagctgacccgccactctcacatttgggctc 155
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 811 AGCACCAGTGCCAAATTTGGTGGTGTACACTAAGTACCAGCAATCTGCTTTGATGGCTCTCC 870
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 96 agcaccagtggtggtcttcagtgctggtgcccagctgacccgccactctcacatttgggctc 155
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 871 AGCACTGGTGTGGCTTTGGTGGTGTCTCAACACCAGTGCCAGCCTTGGCAGTGTGCTC 930
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 156 ttcgctggccttgggtggagctggtgcccagcacca 189
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 931 AACACCAGTACGGTTTGGTGGTGTATGAGCA 954
```

Search completed: April 29, 2002, 19:55:38

us-09-248-178-59.std.rni

Tue Apr 30 14:17:53 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic --nucleic search, using sw model
Run on: April 29, 2002, 19:34:45 ; Search time 8020.64 Seconds
(without alignments)
996.671 Million cell updates/sec

Title: US-09-248-178-59
Perfect score: 382
Sequence: 1 agcgggacgaagctaa.....attaaaaa..... 382

Setting table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797556 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_ov.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: gb_vl.*
16: em_ba.*
17: em_fun.*
18: em_in.*
19: em_mu.*
20: em_on.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description

1	380	99.5	1986	9	HSA293618
2	378.8	99.2	2064	9	HSU92544
3	378.8	99.2	2077	9	BC000304
4	375.6	98.3	1462	6	AX015361
5	369.2	96.6	446	6	AX337185
6	367.4	96.2	384	6	AX106402
7	367.4	96.2	384	6	AX140693
8	367.4	96.2	384	6	AX200553
9	367.4	96.2	384	6	AX267209
10	354.8	92.9	383	6	AX106301
11	354.8	92.9	383	6	AX140592
12	354.8	92.9	383	6	AX200452
13	354.8	92.9	383	6	AX267108
14	338	88.5	2001	9	AF128527
15	335.6	87.9	499	6	AX106292
16	335.6	87.9	499	6	AX140583
17	335.6	87.9	499	6	AX200443
18	335.6	87.9	499	6	AX267099
19	315.8	82.7	1982	9	AF128528
20	241	63.1	1904	9	AF126181
21	237.4	62.1	7781	9	HS1409
22	213.4	55.9	2094	10	MMU277114
23	197.8	51.8	1993	10	RNO293617
24	134.6	35.2	183	11	G06147
25	119	31.2	1785	10	AF319976
26	66.6	17.4	686	8	CNS01D1K
27	66.6	17.4	696	8	CNS01A92
28	58.4	15.3	191923	9	AC087886
29	57.4	15.0	163983	9	AC005519
30	56.4	14.8	270365	1	AF002562
31	55.6	14.6	44776	9	AL138722
32	55.2	14.5	1664	9	HSB01496
33	55.2	14.5	2745	9	AF258554
34	55	14.4	227757	2	AC020706
35	54.8	14.3	57049	9	HS085A12
36	54.8	14.3	163872	2	AC078907
37	54.8	14.3	176291	9	AC024610
38	54.8	14.3	181439	2	AC090611
39	54.6	14.3	201	3	S53307
40	54.6	14.3	2465	6	A31598
41	54.6	14.3	2465	10	MMURFPER
42	54.6	14.3	2465	10	MUSPER
43	54.6	14.3	2711	6	AX099290
44	54.6	14.3	2711	6	AX099294
45	54.6	14.3	2711	9	AF217963

ALIGNMENTS

RESULT 1	HSA293618	1986 bp	mrna	linear	PRI 21-DEC-2000
LOCUS	HSA293618				
DEFINITION	Homo sapiens mRNA for hypothetical protein 11B6, clone XP11B6.				
ACCESSION	AJ293618				
VERSION	AJ293618.1	GI:11967745			
KEYWORDS	HUM11B6 gene; hypothetical protein 11B6.				
SCUPE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 1986)				
JOURNAL	Langnaese, K., Kloos, D.U., Wehnert, M. and Wieacker, P.				
REFERENCE	Cloning a novel protein with similarity to the MAGE family				
AUTHORS	2 (bases 1 to 1986)				
TITLE	Langnaese, K.				
JOURNAL	Direct Submission				
	Submitted (19-JUL-2000) Langnaese K., Institute for Human Genetics,				
	Otto-von-Guericke University, Leipziger Strasse 44, D-39120				
	Magdeburg, D-39120, GERMANY				
FEATURES	Location/Qualifiers				
source	1..1986				

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Xp11B6"
/clone_lib="placenta"
13..1833
CDS
/codon_start=1
/product="hypothetical protein 11B6"
/protein_id="CAC12410.1"
/db_xref="GI:11967746"
/translation="MSDISESGAGLRFQAEASEKSSMMOTLLITVQNVPEIPK
ASKALEVEDKVSQKSGVSKATEVSKIPAREAPATQASSTIOLITDQVLAENKSL
AADTKQNDPQAVTMPATETKKVSHVADTKVNTKAOETEAAPSOAPADEPEPEASAA
QSOENQDTPRKVKAKKARKVHLDEEDSSQASGTTGGRVSKALMASMARRAS
PGPIAFWARRSRTRLAWARALLSLSPKARRGKARRRAKLOSQEPAPPRDV
ALLOGRANDLVKYLAKDOTKIPKRSMDLKDIIKEVDVYPIEIERAGYSLEKVFGI
QLKEIDNDHLXILLSTLEPDAIGLITIKDSPKGLGLMLVLLSTIFMNGNRSSEAVI
EVLRLKLPGRGIIHSLFGDVKKLITDEFVKQYLDYARVPSNPPEYEFFWGLRSYIE
TSKMKVLKFAKQVKQKPEWAAOYREAMEADLKAEEAAEAKAKAEIRAMGILG
SENAAGPCWDEADIGPWAKRIOAGAEAKAEOESGASSTGASTNNNSASASSTS
GGFSAGASLTATLIFGLFAGLGGAGASTSGSSGACGFSYK"
BASE COUNT 528 a 507 c 553 g 398 t
OPIGIN
Query Match 99.5%; Score 380; DB 9; Length 1986;
Best Local Similarity 100.0%; Pred. No. 1.7e-99;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy: 1 agcggagcagaagctaaagccaaagcgaagagtgagcagtcgacgactgtgcca 60
Db 1607 AGGCGGGAGCAGAGCTAAAGCCAAAGCCAAAGAGAGTGGCAGTCCACAC 1656
Qy 61 gtaccagtaccataacagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcg 120
Db 1667 GTACCACTACCAATACAGTGGCAGTGGCAGTGGCAGTGGCAGTGGCAGTGG 1726
Qy 121 gtccagcctgacgcagcactctcacatttgggctcttcagtcgagtcgagtcgagtcg 180
Db 1727 GTGCGAGCCTGACGGCCACTCTCACATTTGGGCTCTTCGCTGGCTTGGTGGTGG 1785
Qy 181 ccagcagcagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcg 240
Db 1787 CCAGCAGCAGTGGCAGTCTGGTGGCTGTGGTGGCTGTGGTGGCTGTGGTGG 1845
Qy 241 gttaatctgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcg 300
Db 1847 GTTAATCTGCCAGTCTTCTCTCAAGCCAGGGTGGTGGTGGTGGTGGTGGTGG 1906
Qy 301 agcactctagcagcagcactatcaatcaattgaagtcgagtcgagtcgagtcgagtcgagtcg 360
Db 1907 AGCACTCTAGCAGCAGCAGTATCAATCAATGAAGTTGACACTCTGCATTAATCTATTG 1966
Qy 361 ccattaaaaaataaaaaa 380
Db 1967 CCATTAAAAAATAAAAAA 1986
RESULT 2
HSU92544 HSU92544 2064 bp mRNA linear PRI 05-JAN-1999
DEFINITION Human hepatocellular carcinoma associated protein (JCL-1) mRNA,
complete cds.
ACCESSION U92544
VERSION U92544.1 GI:4099968
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 2064)
AUTHORS Jin,C.L., Wang,D.Y., Wan,D.F. and Gu,J.R.
TITLE Hepatocellular Carcinoma Associated Gene JCL-1
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2064)
AUTHORS Gu,J.R.
Direct Submission
Submitted (10-MAR-1997) National Laboratory for Oncogenes and
Related Genes, Shanghai Cancer Institute, No. 25, Ln. 2200, Xie-Tu
Rd., Shanghai 200032, People's Republic of China
FEATURES
Location/Qualifiers
1..2064
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..2064
/gene="JCL-1"
67..1897
/gene="JCL-1"
/codon_start=1
/product="hepatocellular carcinoma associated protein"
/protein_id="AAC00728.1"
/db_xref="GI:4099969"
/translation="MSDISESGAGLRFQAEASEKSSMMOTLLITVQNVPEIPK
ASKALFVEDKVSQKSGVSKATEVSKIPAREAPATQASSTIOLITDQVLAENKSL
AADTKQNDPQAVTMPATETKKVSHVADTKVNTKAOETEAAPSOAPADEPEPEASAA
QSOENQDTPRKVKAKKARKVHLDEEDSSQASGTTGGRVSKALMASMARRAS
RGPIAFWARRSRTRLAWARALLSLSPKARRGKARRRAKLOSQEPAPPRDV
ALLOGRANDLVKYLAKDOTKIPKRSMDLKDIIKEVDVYPIEIERAGYSLEKVFGI
QLKEIDNDHLXILLSTLEPDAIGLITIKDSPKGLGLMLVLLSTIFMNGNRSSEAVI
EVLRLKLPGRGIIHSLFGDVKKLITDEFVKQYLDYARVPSNPPEYEFFWGLRSYIE
TSKMKVLKFAKQVKQKPEWAAOYREAMEADLKAEEAAEAKAKAEIRAMGILG
SENAAGPCWDEADIGPWAKRIOAGAEAKAEOESGASSTGASTNNNSASASSTS
GGFSAGASLTATLIFGLFAGLGGAGASTSGSSGACGFSYK"
BASE COUNT 564 a 523 c 570 g 407 t
ORIGIN
Query Match 99.2%; Score 378.8; DB 9; Length 2064;
Best Local Similarity 99.5%; Pred. No. 3.7e-99;
Matches 380; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy: 1 agcggagcagaagctaaagccaaagcgaagagtgagcagtcgacgactgtgcca 60
Db 1661 AGGCGGGAGCAGAGCTAAAGCCAAAGCCAAAGAGAGTGGCAGTCCACAC 1720
Qy 61 gtaccagtaccataacagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcg 120
Db 1721 GTACCACTACCAATACAGTGGCAGTGGCAGTGGCAGTGGCAGTGGCAGTGG 1780
Qy 121 gtccagcctgacgcagcactctcacatttgggctcttcagtcgagtcgagtcgagtcgagtcg 180
Db 1781 GTGCCAGCCTGACGCCACTCTCACATTTGGGCTCTTCGCTGGCTTGGTGGAGCTGGTG 1840
Qy 181 ccagcagcagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcg 240
Db 1841 CCAGCAGCAGTGGCAGTCTGGTGGCTGTGGTGGCTGTGGTGGCTGTGGTGG 1900
Qy 241 gttaatctgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcg 300
Db 1901 GTTAATCTGCCAGTCTTCTCTCAAGCCAGGGTGGTGGTGGTGGTGGTGGTGG 1960
Qy 301 agcactctagcagcagcactatcaatcaattgaagtcgagtcgagtcgagtcgagtcgagtcgagtcg 360
Db 1961 AGCACTCTAGCAGCAGCAGTATCAATCAATGAAGTTGACACTCTGCATTAATCTATTG 2020
Qy 361 ccattaaaaaataaaaaa 382
Db 2021 CCATTAAAAAATAAAAAA 2042
RESULT 3
BC000304 BC000304 2077 bp mRNA linear PRI 12-JUL-2001
LOCUS Homo sapiens, hepatocellular carcinoma associated protein: breast
DEFINITION cancer associated gene 1, clone MGC:8386 IMAGE:2820442, mRNA,
complete cds.
ACCESSION BC000304

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PEMARK
COMMENT
FEATURES
source
CDS
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches 380: Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 aggcggagcagaactaaagccaaagcgaagagtgccagtgccagcactggtacca 60

BC000304.1 GI:12653076
MGC.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2077)
Direct Submission
Submitted (15-NOV-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Prepared by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nsl.nih.gov
Shevchenko.Y., Wetherby.K.D., Beckstrom-Sternberg.S.M.,
Benjamin.B., Blakesley.R.W., Bouffard.G.G., Brinkley.C., Brooks.S.,
Dietrich.N.L., Guan.X., Gupta.J., Ho.S.-L., Karlins.E., Ledespi.R.,
Lim.W., Maduro.O.L., Masello.C., Mastrian.S.D., McCloskey.J.C.,
McDowell.J., Pearson.R., Snyder.B., Slantripop.S., Thomas.P.J.,
Tiongson.E.E., Touchman.J.W., Tsurgeon.C., Vogt.J.L., Walker.M.A.,
Zhang.L.-H. and Green.E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 1 Row: c Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 3288022.
Location/Qualifiers
1..2077
/organism="Homo sapiens"
/db_xref="locusid:10916"
/db_xref="taxon:9606"
/clone="MGC:8396 IMAGE:2820442"
/tissue_type="Lung, small cell carcinoma"
/clone_lib="NIH_MGC_7"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
97..1917
/codon_start=1
/product="hepatocellular carcinoma associated protein;
breast cancer associated gene 1"
/protein_id="AAH00304.1"
/db_xref="GI:12653077"
/translation="MSDTSESGAGLTFQAEASEKSDSSMMOTLLTVTONVVEPIPK
ASKALVSDQVKSAGSVKATIEVSKTPEAREPAOTASSTIOTDVOAEAKNSL
AADTKONADQVATPATETKRVSHVADTKVHTKAQETEAAPSOAPADEPEESAAA
QSEONDTPKVKAKKRVKHLDEGSDSDSOASGTTGGRRVSKALMASMPBPAS
RGIAFWARRASTRLAARWALLSLSPKARRGKARPPAAKLOSSOEAPPPRDV
ALLOGRADNRKYLAKDQTKPIKRSMLKDIKEYIDVPEILIFRAGYSLEKVEGI
OLKEIDKNRHLIYHLSFGVKKLITDEFVKQKLYDYARVPNSHPPTFEFWRYSYE
EVLKRLGFRPKVQKQKPEWAKQYREAMEADLKAAAEAAAEAKARAEIARWGIIGL
TSKMKVLKFPKQVQKQKPEWAKQYREAMEADLKAAAEAAAEAKARAEIARWGIIGL
SENAAGPCNDEADIGPWAKRQIQAQAEAKAKAQSASIGASIGASTINNSASASASTS
GGTFAGASLTLTIFGLFAGLGAGASTSGSSGACGFSYK"
551 a 527 c 587 g 412 t
BASE COUNT
ORIGIN
Query Match 99.2%; Score 378.8; DB 9; Length 2077;
Best Local Similarity 99.5%; Pred. No. 3.7e-99;
Matches 380: Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 aggcggagcagaactaaagccaaagcgaagagtgccagtgccagcactggtacca 60

Db 1691 AGGCGGAGCAGCAAGCTAAAGCCAAAGCCAAAGAGTGGCAGTCCAGCACTGGTGCCA 1750
Qy 61 gtaccagtaccataaacaagtcgaagtcgaagtcgaagtcgaagtcgaagtcgaagtcg 120
Db 1751 GTACCAGTACCATAACAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 1810
Qy 121 gtgcagcctgaccgcactctcacatttgggctcttgcgtggccttggtagagtcggtg 180
Db 1811 GTGCCAGCCTGACCGCCACTCTCACATTITGGGCTCTTCGCTGGCTTGGTGAGCTGGT 1870
Qy 181 ccagccagtcgagcagtcctggtgctggttcttcctcacagtgagaatttagattt 240
Db 1871 CCAGCACCAAGTGGCAGCTCTGGTGGCTTGGTGGCTTGGTGGCTTGGTGGCTTGGT 1930
Qy 241 gttcaatcctaccagtccttctctcaagccaggtgctcctcctcaagaaacctactcaaac 300
Db 1931 GTTAATCTGCCAGTCTTCTCTTCAAGCCAGGTCGATCCTCAGAAACCTACTCAACAC 1990
Qy 301 agcactctagcagcagcactatcaatcaattgaagtcgacactctgattaaatctattt 360
Db 1991 AGCACTCTAGGCAGCCACTATCAATCAATTGAAGTTGACACTCTGCATTAAATCTATT 2050
Qy 361 ccattaaaaaataaaaaa 382
Db 2051 CCATTTCAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 2072
RESULT 4
AX015361 AX015361 1462 bp DNA linear PAT 07-SEP-2000
LOCUS Sequence 20 from Patent WO99511727.
ACCESSION AX015361
VERSION AX015361.1 GI:10041341
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1462)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pillarsky,C.
TITLE Human nucleic acid sequences of normal ovary tissue
JOURNAL Patent: WO 9951727-A 20 14-OCT-1999;
SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILLARSKY CHRISTIAN (DE)
FEATURES Location/Qualifiers
source 1..1462
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 377 a 343 c 399 g 343 t
ORIGIN
Query Match 98.3%; Score 375.6; DB 6; Length 1462;
Best Local Similarity 99.0%; Pred. No. 3e-98;
Matches 378: Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 agcgggagcagaagcctaaagccaaagcgaagagtgccagtcgagcactggtgcca 60
Db 1055 AGCGGGAGCAGCAAGCTAAAGCCAAAGCCAAAGAGTGGCAGTGCAGCACTGGTGCCA 1114
Qy 61 gtaccagtaccataaacaagtcgaagtcgaagtcgaagtcgaagtcgaagtcggtg 120
Db 1115 GTACCAGTACCATAACAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 1174
Qy 121 gtgcagcctgaccgcactctcacatttgggctcttgcgtggccttggtagagtcggtg 180
Db 1175 GTGCCAGCCTGACCGCCACTCTCACATTITGGGCTCTTCGCTGGCTTGGTGAGCTGGT 1234
Qy 181 ccagccagtcgagcagtcctggtgctggttcttcctcacagtgagaatttagattt 240

TITLE Skeiky,Y.A. and Wang,A.
Compositions and methods for the therapy and diagnosis of prostate cancer

JOURNAL Patent: WO 0134802-A 183 17-MAY-2001;
CORIXA CORPORATION (US)

FEATURES Location/Qualifiers
source 1..384

BASE COUNT 102 a 102 c 90 g 89 t 1 others

ORIGIN

Query Match 95.2%; Score 367.4; DB 6; Length 384;
Best Local Similarity 99.0%; Pred. No. 6.le-96;

Matches 379; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Oy 1 agcgaggagcagaagctaaagccaaagcccaag-agagtggcagtgccagcactgggtgcc 59

Db 1 AGCGGGGAGCAGAAAGCTAAAGCCAAAGCCCAAGAGAGTGGCAGTGCCAGCAGCTGGTGC 60

Oy 60 agtaccagttaccataaacagtgccagtgccagtgccagtcagtgcttcagtgct 119

Db 61 AGTACCAGTACCAATAACAGTGCCAGTGCCAGTGCCAGTGCCAGTGCCAGTGCCAGTGCC 120

Oy 120 ggtgccagctgacccgacactctcacatttgggctcttcgctggccttgggtggagctgct 179

Db 121 GGTGCCAGCTGACCGCAGCTCTCACATTTGGGCTCTTCGCTGGCTTGGTGGAGCTGT 180

Oy 180 gccagcacttagcagcactctcaatttgggctcttcgctggccttgggtggagctgct 239

Db 181 GCCAGCAGTACCAATAACAGTGCCAGTGCCAGTGCCAGTGCCAGTGCCAGTGCCAGTGCC 240

Oy 240 ttttaactctgccagctcttcttcaagccagtgctgctcctcagaaacctactcaaca 299

Db 241 TGTTAACTCTGCCAGTCTTCTTCAAGCCAGGCTGATCTCTCAGAAACCTACTCAACA 300

Oy 300 cagcacttagcagcactctcaatttgggctcttcgctggccttgggtggagctgct 359

Db 301 CAGCAGCTCTAGCGAGCAGTCTCAATCAATTTGAAGTTGACACTCTGCATTTATATCTATT 360

Oy 360 gccattataaaaaa 382

Db 361 GCCATTTCAAAAAA 383

RESULT 8
AX200553 384 bp DNA linear PAT 29-AUG-2001

LOCUS Sequence 183 from Patent WO0151633.

DEFINITION AX200553

ACCESSION AX200553

VERSION AX200553.1 GI:15390369

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 384)

Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,

Reed,S.G., Kalos,M.D., Fanger,G.R., Day,C.H., Retter,M.W.,

Stolk,J.A., Skeiky,Y.A., Wang,A., and Meagher,M.J.

Compositions and methods for the therapy and diagnosis of prostate cancer

Patient: WO 0151633-A 183 19-JUL-2001;

CORIXA CORPORATION (US)

Location/Qualifiers

source 1..384

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 102 a 102 c 90 g 89 t 1 others

ORIGIN

Query Match 96.2%; Score 367.4; DB 6; Length 384;
Best Local Similarity 99.0%; Pred. No. 6.le-96;

Matches 379; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Oy 1 agcgaggagcagaagctaaagccaaagcccaag-agagtggcagtgccagcactgggtgcc 59

Db 1 AGCGGGGAGCAGAAAGCTAAAGCCAAAGCCCAAGAGAGTGGCAGTGCCAGCAGCTGGTGC 60

Oy 60 agtaccagttaccataaacagtgccagtgccagtgccagtcagtgcttcagtgct 119

Db 61 AGTACCAGTACCAATAACAGTGCCAGTGCCAGTGCCAGTGCCAGTGCCAGTGCCAGTGCC 120

Oy 120 ggtgccagctgacccgacactctcacatttgggctcttcgctggccttgggtggagctgct 179

Db 121 GGTGCCAGCTGACCGCAGCTCTCACATTTGGGCTCTTCGCTGGCTTGGTGGAGCTGT 180

Oy 180 gccagcacttagcagcactctcaatttgggctcttcgctggccttgggtggagctgct 239

Db 181 GCCAGCAGTACCAATAACAGTGCCAGTGCCAGTGCCAGTGCCAGTGCCAGTGCCAGTGCC 240

Oy 240 ttttaactctgccagctcttcttcaagccagtgctgctcctcagaaacctactcaaca 299

Db 241 TGTTAACTCTGCCAGTCTTCTTCAAGCCAGGCTGATCTCTCAGAAACCTACTCAACA 300

Oy 300 cagcacttagcagcactctcaatttgggctcttcgctggccttgggtggagctgct 359

Db 301 CAGCAGCTCTAGCGAGCAGTCTCAATCAATTTGAAGTTGACACTCTGCATTTATATCTATT 360

Oy 360 gccattataaaaaa 382

Db 361 GCCATTTCAAAAAA 383

RESULT 9
AX267209 384 bp DNA linear PAT 26-OCT-2001

LOCUS Sequence 183 from Patent WO0173032.

DEFINITION AX267209

ACCESSION AX267209

VERSION AX267209.1 GI:16515987

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (sites)

Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,

Kalos,M.D., Fanger,G.R., Retter,M.W., Stolk,J.A., Day,C.H.,

Vedvick,T.S., Carter,D., Li,S.X., Wang,A., Skeiky,Y.A., Hepler,W.I.

and Henderson,R.A.

Compositions and methods for the therapy and diagnosis of prostate cancer

Patent: WO 0173032-A 183 04-OCT-2001;

CORIXA CORPORATION (US)

Location/Qualifiers

source 1..384

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 102 a 102 c 90 g 89 t 1 others

ORIGIN

Query Match 96.2%; Score 367.4; DB 6; Length 384;
Best Local Similarity 99.0%; Pred. No. 6.le-96;

Matches 379; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Oy 1 agcgaggagcagaagctaaagccaaagcccaag-agagtggcagtgccagcactgggtgcc 59

Db 1 AGCGGGGAGCAGAAAGCTAAAGCCAAAGCCCAAGAGAGTGGCAGTGCCAGCAGCTGGTGC 60

Oy 60 agtaccagttaccataaacagtgccagtgccagtgccagtcagtgcttcagtgct 119

Db 61 AGTACCAGTACCAATAACAGTGCCAGTGCCAGTGCCAGTGCCAGTGCCAGTGCCAGTGCC 120

Oy 120 ggtgccagctgacccgacactctcacatttgggctcttcgctggccttgggtggagctgct 179

Db 121 GGTGCCAGCTGACCGCAGCTCTCACATTTGGGCTCTTCGCTGGCTTGGTGGAGCTGT 180

Oy 180 gccagcacttagcagcactctcaatttgggctcttcgctggccttgggtggagctgct 239

Db 181 GCCAGCAGTACCAATAACAGTGCCAGTGCCAGTGCCAGTGCCAGTGCCAGTGCCAGTGCC 240

Oy 240 ttttaactctgccagctcttcttcaagccagtgctgctcctcagaaacctactcaaca 299

Db 241 TGTTAACTCTGCCAGTCTTCTTCAAGCCAGGCTGATCTCTCAGAAACCTACTCAACA 300

Oy 300 cagcacttagcagcactctcaatttgggctcttcgctggccttgggtggagctgct 359

Db 301 CAGCAGCTCTAGCGAGCAGTCTCAATCAATTTGAAGTTGACACTCTGCATTTATATCTATT 360

Oy 360 gccattataaaaaa 382

Db 361 GCCATTTCAAAAAA 383

RESULT 9
AX267209 384 bp DNA linear PAT 26-OCT-2001

LOCUS Sequence 183 from Patent WO0173032.

DEFINITION AX267209

ACCESSION AX267209

VERSION AX267209.1 GI:16515987

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (sites)

Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,

Kalos,M.D., Fanger,G.R., Retter,M.W., Stolk,J.A., Day,C.H.,

Vedvick,T.S., Carter,D., Li,S.X., Wang,A., Skeiky,Y.A., Hepler,W.I.

and Henderson,R.A.

Compositions and methods for the therapy and diagnosis of prostate cancer

Patent: WO 0173032-A 183 04-OCT-2001;

CORIXA CORPORATION (US)

Location/Qualifiers

source 1..384

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 102 a 102 c 90 g 89 t 1 others

ORIGIN

```

Oy 120 ggtgccagcctgaaccgacactctcacatttggtgctctcgctggccttggtgagctgct 179
Db 121 GGTGCCAGCCTGACCGCCACTCTCACATTGGGCTCTTCGCTGGGCTTTGGTGGAGCTGGT 180
Oy 180 gccagcaccagtgagcagctctggtgctctggtttctctcacaagtgcagattttagatat 239
Db 181 GCCAGCACAGTGGCAGCTCTGGTGGCTTCTCTCTACAAAGTGAGATTTAGATAT 240
Oy 240 ttttaattcctgccagctcttctctcaagccaggggtgcacatctcagaacctactcaaca 299
Db 241 TGTTAATCTGCCAGCTCTTCTCTCAAGCCAGGGTGCAATCTCTCAAGAACCTTACTCAACA 300
Oy 300 cagcactctaggcagcactatcaatcaattgaattgcacactctgcattaaatctattt 359
Db 301 CAGCCTCTAGGAGCCACTCAATCAATGAATTGACACTCTGCATTATATCTATT 360
Oy 360 gccattaaaaaahhaaahhaa 382
Db 361 GCCATTTCAAAAAAAAAAAAAA 382

RESULT 10
LOCUS AX106301 383 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 82 from Patent WO0125272.
ACCESSION AX106301
VERSION AX106301.1 GI:13921987
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 383)
AUTHORS Xu,J., Skeiky,Y.A., Reed,S.G. and Cheever,M.A.
TITLE Compositions and methods for therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0125272-A 82 12-APR-2001;
CORIXA CORPORATION (US)
FEATURES
Location/Qualifiers
source
misc_feature 1..383
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 102 a 102 c 89 g 89 t 1 others
ORIGIN

Query Match 92.9%; Score 354.8; DB 5; Length 383;
Best Local Similarity 98.7%; Pred. No. 2.8e-92;
Matches 378; Conservative 0; Mismatches 3; Indels 2; Gaps 2:

Oy 1 agcgggagcagaagcctaaagccaaagcccaag-aggatggcagtgccagcactgagtc 59
Db 1 AGCGGGAGCAGAAGCTAAAGCCAAAGCCAAAGAGTGGCAGTGCCAGCACTGGTGCC 50
Oy 60 agtaccagctaccataaacagtgccagtgccagtcagtcagtcagtcagtcagtcagtc 119
Db 61 AGTACCAGTACCAATAACA-TGCCAGTGCAGTGGCAGCAGCAGTGGTGGCTTACAGTCT 119
Oy 120 ggtgccagcctgaccgacactctcacatttggtgctctcgctggccttggtgagctgct 179
Db 120 GGTGCCAGCCTGACCGCCACTCTCACATTGGGCTCTTCGCTGGCCTTGGTGGAGCTGGT 179
Oy 180 gccagcaccagtgagcagctctggtgctctggtttctctcacaagtgcagattttagatat 239
Db 180 GCCAGCACAGTGGCAGCTCTGGTGGCTTCTCTCTACAAAGTGAGATTTAGATAT 239
Oy 240 ttttaattcctgccagctcttctctcaagccaggggtgcacatctcagaacctactcaaca 299
Db 240 TGTTAATCTGCCAGCTCTTCTCTCAAGCCAGGGTGCAATCTCTCAAGAACCTTACTCAACA 299
Oy 300 cagcactctaggcagcactatcaatcaattgaattgcacactctgcattaaatctattt 359
Db 300 CAGCCTCTAGGAGCCACTCAATCAATGAATTGACACTCTGCATTATATCTATT 359
Oy 360 gccattaaaaaahhaaahhaa 382
Db 360 GCCATTTCAAAAAAAAAAAAAA 382

RESULT 12
LOCUS AX200452 383 bp DNA linear PAT 29-AUG-2001
DEFINITION Sequence 82 from Patent WO0151633.
ACCESSION AX200452

```

```

Db 300 CAGCAGCTCTGGCAGCCACTATCAATTAATTGAAGTTGACACTCTGCATTAAATCTATT 359
Oy 360 gccattaaaaaahhaaahhaa 382
Db 360 GCCATTTCAAAAAAAAAAAAAA 382

RESULT 11
LOCUS AX140592 383 bp DNA linear PAT 31-MAY-2001
DEFINITION Sequence 82 from Patent WO0134802.
ACCESSION AX140592
VERSION AX140592.1 GI:14280710
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 383)
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Reed,S.G., Kalos,M.D., Retter,M.W., Stolk,J.A., Day,C.H.,
Skeiky,Y.A. and Wang,A.
TITLE Compositions and methods for the therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0134802-A 82 17-MAY-2001;
CORIXA CORPORATION (US)
FEATURES
Location/Qualifiers
source
misc_feature 1..383
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 102 a 102 c 89 g 89 t 1 others
ORIGIN

Query Match 92.9%; Score 354.8; DB 6; Length 383;
Best Local Similarity 98.7%; Pred. No. 2.8e-92;
Matches 378; Conservative 0; Mismatches 3; Indels 2; Gaps 2:

Oy 1 agcgggagcagaagcctaaagccaaagcccaag-aggatggcagtgccagcactgagtc 59
Db 1 AGCGGGAGCAGAAGCTAAAGCCAAAGCCAAAGAGTGGCAGTGCCAGCACTGGTGCC 50
Oy 60 agtaccagctaccataaacagtgccagtgccagtcagtcagtcagtcagtcagtcagtc 119
Db 61 AGTACCAGTACCAATAACA-TGCCAGTGCAGTGGCAGCAGCAGTGGTGGCTTACAGTCT 119
Oy 120 ggtgccagcctgaccgacactctcacatttggtgctctcgctggccttggtgagctgct 179
Db 120 GGTGCCAGCCTGACCGCCACTCTCACATTGGGCTCTTCGCTGGCCTTGGTGGAGCTGGT 179
Oy 180 gccagcaccagtgagcagctctggtgctctggtttctctcacaagtgcagattttagatat 239
Db 180 GCCAGCACAGTGGCAGCTCTGGTGGCTTCTCTCTACAAAGTGAGATTTAGATAT 239
Oy 240 ttttaattcctgccagctcttctctcaagccaggggtgcacatctcagaacctactcaaca 299
Db 240 TGTTAATCTGCCAGCTCTTCTCTCAAGCCAGGGTGCAATCTCTCAAGAACCTTACTCAACA 299
Oy 300 cagcactctaggcagcactatcaatcaattgaattgcacactctgcattaaatctattt 359
Db 300 CAGCCTCTAGGAGCCACTCAATCAATGAATTGACACTCTGCATTATATCTATT 359
Oy 360 gccattaaaaaahhaaahhaa 382
Db 360 GCCATTTCAAAAAAAAAAAAAA 382

RESULT 12
LOCUS AX200452 383 bp DNA linear PAT 29-AUG-2001
DEFINITION Sequence 82 from Patent WO0151633.
ACCESSION AX200452

```

```

VERSION      AX200452.1  GI:15390261
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 383)
              Xu, J., Dillon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
              Peed, S.G., Kalos, M.D., Fanger, G.R., Day, C.H., Retter, M.W.,
              Stolk, J.A., Skelky, Y.A., Wang, A., and Meagher, M.J.
              Stolk, J.A., Skelky, Y.A., Wang, A., and Meagher, M.J.
              Compositions and methods for the therapy and diagnosis of prostate
              cancer
              Patent: WO 0151633-A 82 19-JUL-2001;
              CORIXA CORPORATION (US)
              Location/Qualifiers
              1..383
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              102 a 102 c 89 g 89 t 1 others
              BASE COUNT
              ORIGIN

Query Match      92.9%; Score 354.8; DB 5; Length 383;
Best Local Similarity 98.7%; Pred. No. 2.8e-92;
Matches 378; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1 agcgggagcagaagctaaagccaaagccaaag-agaagtgccagtgccagcactgagtgcc 59
Db 1 AGCGGGAGCAGAGCTAAAGCCAAAGCCAAAGAGAGTGGCAGTGGCCACGACGTCGTCGCC 60

QY 60 agtaccagtcaccataaagcagtgccagtgccagtcagtcagtcagtcagtcagtcagtc 119
Db 61 AGTACCAGTACCAATAACA-TGCCAGTGCCAGTGCCAGCAGCAGTGCGTTCAGTGCT 119

QY 120 ggtgcagcctgaccgcccactctcacatttgggctctcgtgccccttggaggcctggt 179
Db 120 GGTGCCAGCCIGACCGCCACTCTCACATTGGGCTCTTGGCTGGCTTGGTGAGCTGCT 179

QY 180 gcagcaccagtcagcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 239
Db 180 GCCAGCAGCAGTGGCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 239

QY 240 tggtaactctgcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 299
Db 240 TGTTAATCTGCCAGCTCTTCTTCTTCAAGCCAGGCTGCTTCTTCAAGCCAGGCTGCTTCT 299

QY 300 gcagcactctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 359
Db 300 CAGCAGCTCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 359

QY 360 gccattaaaaaataaaaaa 382
Db 360 GCCATTTCAAAAAATAAAAAA 382

RESULT 13
LOCUS      AX267108
DEFINITION Sequence 82 from Patent WO0173032.
ACCESSION AX267108
VERSION    AX267108.1  GI:15515893
KEYWORDS
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
              1 (sites)
              Xu, J., Dillon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
              Kalos, M.D., Fanger, G.R., Retter, M.W., Stolk, J.A., Day, C.H.,
              Vedvick, I.S., Carter, D., Li, S.X., Wang, A., Skelky, Y.A., Hepler, W.T.
              and Henderson, R.A.
              Compositions and methods for the therapy and diagnosis of prostate
              cancer

TITLE

```

```

JOURNAL      Patent: WO 0173032-A 82 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
SOURCE       Location/Qualifiers
              1..383
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              102 a 102 c 89 g 89 t 1 others
              BASE COUNT
              ORIGIN

Query Match      92.9%; Score 354.8; DB 6; Length 383;
Best Local Similarity 98.7%; Pred. No. 2.8e-92;
Matches 378; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1 agcgggagcagaagctaaagccaaagccaaag-agaagtgccagtgccagcactgagtgcc 59
Db 1 AGCGGGAGCAGAGCTAAAGCCAAAGCCAAAGAGAGTGGCAGTGGCCACGACGTCGTCGCC 60

QY 60 agtaccagtcaccataaagcagtgccagtgccagtcagtcagtcagtcagtcagtcagtc 119
Db 61 AGTACCAGTACCAATAACA-TGCCAGTGCCAGTGCCAGCAGCAGTGCGTTCAGTGCT 119

QY 120 ggtgcagcctgaccgcccactctcacatttgggctctcgtgccccttggaggcctggt 179
Db 120 GGTGCCAGCCIGACCGCCACTCTCACATTGGGCTCTTGGCTGGCTTGGTGAGCTGCT 179

QY 180 gcagcaccagtcagcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 239
Db 180 GCCAGCAGCAGTGGCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 239

QY 240 tggtaactctgcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 299
Db 240 TGTTAATCTGCCAGCTCTTCTTCTTCAAGCCAGGCTGCTTCTTCAAGCCAGGCTGCTTCT 299

QY 300 gcagcactctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 359
Db 300 CAGCAGCTCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 359

QY 360 gccattaaaaaataaaaaa 382
Db 360 GCCATTTCAAAAAATAAAAAA 382

RESULT 14
LOCUS      AF128527
DEFINITION Homo sapiens clone sn1 breast cancer associated gene 1 protein
ACCESSION AF128527
VERSION    AF128527.1  GI:4928043
KEYWORDS   human.
SOURCE     Homo sapiens
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 2001)
              Kurt, R.A. and Urban, W.J.
              Isolation of breast cancer associated gene 1 from normal mammary
              epithelium
              unpublished
              2 (bases 1 to 2001)
              Kurt, R.A. and Urban, W.J.
              Direct Submission
              Submitted (15-FEB-1999) Cellular Immunology, Earle A. Childs
              Research Institute, Franz Cancer Research Center, 4805 NE Glisan,
              Portland, OR 97213, USA
              Location/Qualifiers
              1..2001
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="sn1"
              /tissue="sn1"
              /tissue_type="normal mammary epithelium"
              1..2001
              gene

```

```

/ gene="BCG1"
70..1930
/codon_start=1
/product="breast cancer associated gene 1 protein"
/protein_id="AAD33392.1"
/db_xref="GI:4928044"
/translation="MSDTSESGAGLTRFQAEASEKDSQSSMMQTLLTIVTQVVEPTEPK
ASKALEVEDKVKSKASQKATKTEVSKPEAREKAPATQSSATITQLTIDTQVLAENKSL
AADTTRQNADPOAVTIPATETFKVSHVADTKVNTKQETEAAPSOAPADEPEPESAAA
QSQENODTRPKVKAKKARKVHLGDGEDGSDSOASGTTGGRRVSKALMAGMARRPV
RGP1AFWARRARRTLAAWARRALLURSPKARGKARRAAKLOSSPEAPPPRVDI
ALLOQGNADLVYLLAKDQTKIPIKRSDMLKDIIKEIDYVPIEIIERAGYSLEKVG1
ALQGDKNHDLYILLSTLEPTDAGILTTKDSPLGLUMLVLLSITFMNGNRSEAVIN
EVLKGLGFPACKVQFHHSLDGVKKLLITDEFVKOKYLDYRVFNSNPPEYFEFGLRGSLYE
TSKNVYLKFCAGVQKDFKWAQOYREAMEADLKAAAEAAAEAKARAEIIPARMGILGL
SENKAGPCWDEADIPGNAPKIQAGAEAKAKQESGASSTGASITSTHNSASASASTS
GFESGASLTALTITGLFLAGUGGAGASTGSGSGACGFSYK"
BASE COUNT      521 a      515 c      559 g      395 t
ORIGIN

Query Match      98.5%   Score 338   DB 9:   Length 2001:
Best Local Similarity 100.0%   Pred. No. 2.7e-87:
Matches 338:   Conservative 0:   Mismatches 0:   Indels 0:   Gaps 0:

QY 1 aggcggagcgaagcgttaagccaaagccaaagagatggcagctgcacactggatgcc 60
      |||
Db 1664 AGGCGGAGCAGAAAGCTAAAGCCAAAGCCCAAGAGAGTGGCAGTGCACGACTGGTGCCA 1723
      |||

QY 61 gtaccagttaccaataaaccagtcgagtcgccagtcgccagcaccagtggtgcttcagtcgt 120
      |||
Db 1724 GTACCACTACCAATAACAGTCCAGTCCAGTGCACGACAGTGGTGGCTTCAGTGGCTG 1783
      |||

QY 121 gtgcagcgtgaccgccactctcacattgggctcttcgctgacctggctggagagctgggtg 180
      |||
Db 1784 GTGCCAGCCTGACCCGCCACICTCACATITGGGCTCTTCGCTGGCCITGGTGGAGCTGGTG 1843
      |||

QY 181 ccagcaccagtcgagcctctggtgctgctggtttctctcacaagtgagatttagatatt 240
      |||
Db 1844 CCAGCACCAGTGGCAGCTCTGGTGCTGTGTTCTCTCTACAAAGTGAGATTTTAGATATT 1903
      |||

QY 241 gttaatctgcgaagtctttctcttcaagccaggtgcatctctcagaacactactcaacac 300
      |||
Db 1904 GTTAATCTTGCCAGCTTTCITCTCAAGCCAGGGTGTCATCTCTCAGAAACCTACTACAC 1963
      |||

QY 301 agcactctaggcagccactatcaatcaattgaagtga 338
      |||
Db 1964 AGCACTCTAGCAGCACCAGTATCAATCAATTGAAGTTGA 2001
      |||

RESULT 15
AX105292      AX106292      499 bp      DNA      linear      PAT 30-APR-2000
LOCUS      AX105292      Sequence 73 from Patent WO0125272.
DEFINITION      AX105292
ACCESSION      AX106292
VERSION      AX106292.1  GI:13321978
KEYWORDS
SOURCE
human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 499)
XU,J., Skeiky,Y.A., Reed,S.G. and Cheever,M.A.
Compositions and methods for therapy and diagnosis of prostate
cancer
Patent: WO 0125272-A 73 12-APR-2001:
CORIXA CORPORATION (US)
Location/Qualifiers
1..499
/organism="Homo sapiens"
/db_xref="taxon:9606"
misc feature 1..499

```

1000000

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic --nucleic search, using sw model

Run on: April 30, 2002, 10:36:25 ; Search time 308.52 Seconds
(without alignments)
1004.220 Million cell updates/sec

Title: US-09-248-178-61
Perfect score: 1368
Sequence: 1 ccagtgcgcgcgtaatac.....cgcgcttgagtaatacatnn 1368

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 10

Total number of hits satisfying chosen parameters: 107588

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	5.0	703	4	US-08-998-416-136 Sequence 136, Appl
2	68	5.0	716	4	US-08-998-416-55 Sequence 55, Appl
3	68	5.0	793	4	US-09-020-956-33 Sequence 33, Appl
4	68	5.0	793	4	US-09-030-607-33 Sequence 33, Appl
5	68	5.0	801	2	US-08-975-316-50 Sequence 50, Appl
6	68	5.0	814	4	US-09-020-956-1 Sequence 1, Appl
7	68	5.0	814	4	US-09-030-607-1 Sequence 1, Appl
8	68	5.0	815	4	US-09-020-956-2 Sequence 2, Appl
9	68	5.0	815	4	US-09-030-607-2 Sequence 2, Appl
10	68	5.0	949	2	US-08-713-000-4 Sequence 4, Appl
11	68	5.0	949	2	US-08-975-316-4 Sequence 4, Appl
12	68	5.0	949	2	US-09-211-710-4 Sequence 4, Appl
13	68	5.0	2961	4	US-08-446-935-6 Sequence 6, Appl
14	68	5.0	3792	2	US-08-992-334-1 Sequence 1, Appl
15	68	5.0	3792	2	US-08-302-752-1 Sequence 1, Appl
16	68	5.0	4145	3	US-08-651-472-52 Sequence 52, Appl
17	68	5.0	4145	3	US-08-358-928-52 Sequence 52, Appl
18	68	5.0	4277	3	US-08-651-472-63 Sequence 63, Appl
19	68	5.0	4277	3	US-08-358-928-63 Sequence 63, Appl
20	68	5.0	5234	2	US-08-992-334-2 Sequence 2, Appl
21	68	5.0	5234	3	US-08-302-752-2 Sequence 2, Appl
22	68	5.0	5356	4	US-08-446-935-1 Sequence 1, Appl
23	68	5.0	5532	3	US-08-651-472-72 Sequence 72, Appl
24	68	5.0	5532	3	US-08-358-928-72 Sequence 72, Appl
25	68	5.0	6722	2	US-08-992-334-3 Sequence 3, Appl
26	68	5.0	6722	3	US-08-302-752-3 Sequence 3, Appl
27	68	5.0	6811	3	US-08-651-472-67 Sequence 67, Appl

C 28	68	5.0	6811	4	US-08-358-928-67	Sequence 67, Appl
C 29	68	5.0	6926	3	US-08-651-472-69	Sequence 69, Appl
C 30	68	5.0	6926	4	US-08-358-928-69	Sequence 69, Appl
C 31	62	4.5	558	4	US-08-930-285-17	Sequence 17, Appl
C 32	52	4.5	698	4	US-08-998-416-54	Sequence 54, Appl
C 33	62	4.5	713	4	US-08-998-416-135	Sequence 135, Appl
C 34	62	4.5	5534	1	US-08-452-267-3	Sequence 3, Appl
C 35	62	4.5	5534	3	US-09-123-644-3	Sequence 3, Appl
C 36	62	4.5	7659	4	US-09-128-314-4	Sequence 4, Appl
C 37	62	4.5	8157	4	US-09-128-314-3	Sequence 3, Appl
C 38	61	4.5	2363	1	US-08-096-947-2	Sequence 2, Appl
C 39	61	4.5	2363	1	US-07-919-140B-2	Sequence 2, Appl
C 40	61	4.5	2363	4	US-08-916-232-2	Sequence 2, Appl
C 41	61	4.5	2363	5	PCT-US93-06939-2	Sequence 2, Appl
C 42	55	4.0	5618	3	US-08-799-569-1	Sequence 1, Appl
C 43	51	3.7	329	1	US-08-703-809-9	Sequence 9, Appl
C 44	51	3.7	329	1	US-08-703-808-9	Sequence 9, Appl
C 45	51	3.7	329	2	US-08-703-807-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-08-998-416-136
: Sequence 136, Application US/08998416
: Patent No. 6239264
: GENERAL INFORMATION:
: APPLICANT: Philippsen, Peter
: APPLICANT: Pohlmann, Rainer
: APPLICANT: Steiner, Sabine
: APPLICANT: Mohr, Christine
: APPLICANT: Wendland, Jurgen
: APPLICANT: Knechtie, Philipp
: APPLICANT: Rebischung, Corinne
: TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSTYII
: TITLE OF INVENTION: AND USES THEREOF
: NUMBER OF SEQUENCES: 1152
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 6239264artis Corporation
: STREET: 3054 Cornwallis Road
: CITY: Research Triangle Park
: STATE: No. 6239264th Carolina
: COUNTRY: USA
: ZIP: 27709
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/998,416
: FILING DATE: 24-DEC-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: CH 0016/97
: FILING DATE: 31-DEC-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Meigs, J. Timothy
: REGISTRATION NUMBER: 38,241
: REFERENCE/DOCKET NUMBER: PF/5-30306/A/GCG1976
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8587
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 136:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 703 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: PAG1051UP


```

CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/030.607
FILING DATE: 25-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.427C3
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-09-030-607-33

```

```

Query Match          5.0%; Score 68; DB 4; Length 793;
Best Local Similarity 100.0%; Pred. No. 8.6e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1299 gccaccgcgggtgagctccagctttgtccctttagtgagggttaattgcgcgttcgc 1358
|||||
Db 425 gccaccgcgggtgagctccagctttgtccctttagtgagggttaattgcgcgttcgc 484
|||||
Oy 1359 gtaatcat 1365
|||||
Db 485 GTAATCAT 492

```

```

RESULT 5
US-08-975-316-50/c
Sequence 50, Application US/08975316
Patent No. 5952486
GENERAL INFORMATION:
APPLICANT: BLOKSBERG, Leonard N., HAVUKKALA, Ilkka
APPLICANT: and GRIERSON, Alastair W.
TITLE OF INVENTION: MATERIALS AND METHODS FOR
TITLE OF INVENTION: THE MODIFICATION OF PLANT LIGNIN CONTENT
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,316
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: 08/713,000
FILING DATE: September 11, 1996
ATTORNEY/AGENT INFORMATION:
NAME: SLEATH, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000/1003C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 801 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-975-316-50

```

```

Query Match          5.0%; Score 68; DB 2; Length 801;
Best Local Similarity 100.0%; Pred. No. 8.6e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1299 gccaccgcgggtgagctccagctttgtccctttagtgagggttaattgcgcgttcgc 1358
|||||
Db 246 gccaccgcgggtgagctccagctttgtccctttagtgagggttaattgcgcgttcgc 187
|||||
Oy 1359 gtaatcat 1366
|||||
Db 186 GTAATCAT 179

```

```

RESULT 5
US-09-020-956-1
Sequence 1, Application US/09020956
Patent No. 6261562
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillin, David C.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
NUMBER OF SEQUENCES: 178
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/020,956
FILING DATE: 09-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.427C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 814 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-09-020-956-1

```

```
Query Match          5.0%: Score 68; DB 4: Length 814;
Best Local Similarity 100.0%: Pred. No. 8.6e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1299 gccaccgcggtgagctccagcttttgccttttagtgagggttaattgcgcgttggc 1358
|||||
Db 372 GCCACCGCGGTGAGCTCCAGCTTTTGTTCCTTTAGTGAGGGTTAATTGCGCGCTTGGC 431

OY 1359 gtaatcat 1366
|||||
Db 432 GTAATCAT 439

RESULT 7
US-09-030-607-1
: Sequence 1, Application US/09030507
: Patent No. 6262245
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
: NUMBER OF SEQUENCES: 224
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED and BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: FILING DATE: 25-FEB-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Maki, David J.
: REGISTRATION NUMBER: 31,392
: REFERENCE/DOCKET NUMBER: 210121.427C3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 814 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cdna
US-09-030-607-1

Query Match          5.0%: Score 68; DB 4: Length 814;
Best Local Similarity 100.0%: Pred. No. 8.6e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1299 gccaccgcggtgagctccagcttttgccttttagtgagggttaattgcgcgttggc 1358
|||||
Db 474 GCCACCGCGGTGAGCTCCAGCTTTTGTTCCTTTAGTGAGGGTTAATTGCGCGCTTGGC 483

OY 1359 gtaatcat 1366
|||||
Db 484 GTAATCAT 491

RESULT 9
US-09-030-607-2
: Sequence 2, Application US/09030507
: Patent No. 6262245
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
: NUMBER OF SEQUENCES: 224
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED and BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
```

US-08-713-000-4

Query Match 5.0%; Score 68; DB 2; Length 949;
Best Local Similarity 100.0%; Pred. No. 8.4e-18; Indels 0; Gaps 0;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1299 gccaccggtgagctccagcttttgcctttagtgagggttaattgcgcgttggc 1358
|||||
Db 248 GCCACCGGTGGAGCTCCAGCTTTTGTCCCTTAGTGAGGTTAATTGCCGCTTGGC 189
|||||

Oy 1359 gtaatacat 1366
|||||
Db 188 GTAATCAT 181
|||||

RESULT 11
US-08-975-316-4/c
Sequence 4, Application US/08975316
Patent No. 5952486
GENERAL INFORMATION:
APPLICANT: BLOKSBERG, Leonard N., HAVUKKALA, Ilkka
APPLICANT: GRIERSON, Alastair W.
TITLE OF INVENTION: MATERIALS AND METHODS FOR
TITLE OF INVENTION: THE MODIFICATION OF PLANT LIGNIN CONTENT
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,316
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/713,000
FILING DATE: September 11, 1996
ATTORNEY/AGENT INFORMATION:
NAME: SLEATH, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000/1003C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 949 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-975-316-4

Query Match 5.0%; Score 68; DB 2; Length 949;
Best Local Similarity 100.0%; Pred. No. 8.4e-18; Indels 0; Gaps 0;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1299 gccaccggtgagctccagcttttgcctttagtgagggttaattgcgcgttggc 1358
|||||
Db 248 GCCACCGGTGGAGCTCCAGCTTTTGTCCCTTAGTGAGGTTAATTGCCGCTTGGC 189
|||||

Oy 1359 gtaatacat 1366
|||||
Db 188 GTAATCAT 181
|||||

US-09-030-607-2

Query Match 5.0%; Score 68; DB 4; Length 816;
Best Local Similarity 100.0%; Pred. No. 8.6e-18; Indels 0; Gaps 0;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1299 gccaccggtgagctccagcttttgcctttagtgagggttaattgcgcgttggc 1358
|||||
Db 424 GCCACCGGTGGAGCTCCAGCTTTTGTCCCTTAGTGAGGTTAATTGCCGCTTGGC 483
|||||

Oy 1359 gtaatacat 1366
|||||
Db 484 GTAATCAT 491
|||||

RESULT 10
US-08-713-000-4/c
Sequence 4, Application US/08713000
Patent No. 5850020
GENERAL INFORMATION:
APPLICANT: Bloksberg, Leonard N.
APPLICANT: Havukkala, Ilkka
APPLICANT: Grierson, Alastair
TITLE OF INVENTION: MATERIALS AND METHODS FOR THE
TITLE OF INVENTION: MODIFICATION OF PLANT LIGNIN CONTENT
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Speckman Picard PLLC
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release 4.0, Version 4.1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/713,000
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 949 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

RESULT 12
US-09-211-710-4/c
: Sequence 4, Application US/09211710A
: Patent No. 6204434
: GENERAL INFORMATION:
: APPLICANT: Bloksberg, Leonard N.
: APPLICANT: Havukkala, Ilkka
: APPLICANT: Grierson, Alastair
: TITLE OF INVENTION: Materials and Methods for the
: TITLE OF INVENTION: Modification of Plant Lignin Content
: FILE REFERENCE: 11090.1003c3
: CURRENT APPLICATION NUMBER: US/09/211,710A
: CURRENT FILING DATE: 1998-12-14
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 4
: LENGTH: 949
: TYPE: DNA
: ORGANISM: Pinus radiata
: FEATURE:
: NAME/KEY: unsure
: LOCATION: (1)...(949)
: OTHER INFORMATION: n at all occurrences indicates unsure
US-09-211-710-4

Query Match 5.0%; Score 68; DB 4; Length 949;
Best Local Similarity 100.0%; Pred. No. 8.4e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1299 gccaccggtgagctccagcttttgccttagtgagggtaattgcgcgttggc 1358
|||||
Db 248 gccaccggtgagctccagcttttgccttagtgagggtaattgcgcgttggc 189
Oy 1359 gtaatcat 1365
|||||
Db 188 gtaatcat 181

RESULT 13
US-08-446-935-6
: Sequence 6, Application US/08446935
: Patent No. 6187991
: GENERAL INFORMATION:
: APPLICANT: Soeller, Walter C.
: APPLICANT: Carty, Maynard D.
: APPLICANT: Kreutter, David K.
: TITLE OF INVENTION: TRANSGENIC ANIMAL MODELS FOR TYPE II
: TITLE OF INVENTION: DIABETES MELLITUS
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pfizer Inc.
: STREET: 235 East 42nd Street, 20th Floor
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10017-5755
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/446,935
: FILING DATE:
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: Sheyka, Robert F.
: REGISTRATION NUMBER: 31,304
: REFERENCE/DOCKET NUMBER: PC8153
: TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)573-1189
TELEFAX: (212)573-1939
TELEX: N/A
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2961 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: circular
: MOLECULE TYPE: DNA (genomic)
US-08-446-935-6
Query Match 5.0%; Score 68; DB 4; Length 2961;
Best Local Similarity 100.0%; Pred. No. 7.2e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1299 gccaccggtgagctccagcttttgccttagtgagggtaattgcgcgttggc 1358
|||||
Db 743 gccaccggtgagctccagcttttgccttagtgagggtaattgcgcgttggc 802
Oy 1359 gtaatcat 1365
|||||
Db 803 gtaatcat 810
RESULT 14
US-08-992-334-1
: Sequence 1, Application US/08992334
: Patent No. 5919678
: GENERAL INFORMATION:
: APPLICANT: Gruss, Alexandra
: APPLICANT: Maguin, Emmanuelle
: TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
: TITLE OF INVENTION: PLASMID
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Christie Parker & Hale, LLP
: STREET: 350 West Colorado Boulevard, Suite 500
: CITY: Pasadena
: STATE: California
: COUNTRY: United States
: ZIP: 91105
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/992,334
: FILING DATE: 17-DEC-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,752
: FILING DATE: 24-DEC-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/FR93/00248
: FILING DATE: 12-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 992/03034
: FILING DATE: 13-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Prout, D. Bruce
: REGISTRATION NUMBER: 20958
: REFERENCE/DOCKET NUMBER: C93:31779
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (526) 795-9900
: TELEFAX: (626) 577-8800
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3792 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both

Query Match	Best Local Similarity	Score	DB 2:	Length	3792:
Matches 68: Conservative	0: Mismatches	0: Indels	0: Gaps	0:	
OY 1299 gccaccgagtgagctccagcttttgccttttagtgagggttaattgcgccttggc	1358				
DB 3558 gccaccgagtgagctccagcttttgccttttagtgagggttaattgcgccttggc	3517				
OY 1359 gtaatcat 1365					
DB 3618 gtaatcat 3625					
RESULT 15					
US-08-302-752-1					
Sequence 1, Application US/08302752					
Patent No. 6025190					
GENERAL INFORMATION:					
APPLICANT:					
TITLE OF INVENTION: THERMOSENSIBLE PLASMID					
NUMBER OF SEQUENCES: 3					
COMPUTER READABLE FORM:					
MEDIUM TYPE: Floppy disk					
COMPUTER: IBM PC compatible					
OPERATING SYSTEM: PC-DOS/MS-DOS					
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)					
CURRENT APPLICATION DATA:					
APPLICATION NUMBER: US/08/302,752					
FILING DATE:					
CLASSIFICATION: 435					
PRIOR APPLICATION DATA: FR 9203034					
FILING DATE: 13-MAR-1992					
PRIOR APPLICATION NUMBER: WO FR/93/00248					
FILING DATE: 12-MAR-1993					
INFORMATION FOR SEQ ID NO: 1:					
SEQUENCE CHARACTERISTICS:					
LENGTH: 3792 base pairs					
TYPE: nucleic acid					
STRANDEDNESS: single					
TOPOLOGY: linear					
MOLECULE TYPE: DNA (genomic)					
US-08-302-752-1					
Query Match	5.0%	Score 68:	DB 3:	Length 3792:	
Best Local Similarity	100.0%	Pred. No. 7e-18:	0: Mismatches	0: Indels	0: Gaps
Matches 68: Conservative	0: Mismatches	0: Indels	0: Gaps	0:	
OY 1299 gccaccgagtgagctccagcttttgccttttagtgagggttaattgcgccttggc	1358				
DB 3558 gccaccgagtgagctccagcttttgccttttagtgagggttaattgcgccttggc	3517				
OY 1359 gtaatcat 1365					
DB 3618 gtaatcat 3625					
RESULT 16					
US-08-651-472-52/c					
Sequence 62, Application US/08651472					
Patent No. 5103244					
GENERAL INFORMATION:					
APPLICANT:					
TITLE OF INVENTION:					
NUMBER OF SEQUENCES:					
COMPUTER READABLE FORM:					
MEDIUM TYPE:					
COMPUTER:					
OPERATING SYSTEM:					
SOFTWARE:					
CURRENT APPLICATION DATA:					
APPLICATION NUMBER:					
FILING DATE:					
CLASSIFICATION:					
PRIOR APPLICATION DATA:					
FILING DATE:					
PRIOR APPLICATION NUMBER:					
FILING DATE:					
INFORMATION FOR SEQ ID NO: 1:					
SEQUENCE CHARACTERISTICS:					
LENGTH:					
TYPE:					
STRANDEDNESS:					
TOPOLOGY:					
MOLECULE TYPE:					

```

: TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF CHIMERIC
: TITLE OF INVENTION: VIRUSES CONTAINING HUMAN IMMUNODEFICIENCY VIRUS TYPE 1
: TITLE OF INVENTION: (HIV-1) ANTIGENS
: NUMBER OF SEQUENCES: 95
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W., Suite 500
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/358-928
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/914,738
: FILING DATE: 20-JUL-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/750,080
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/156/IMMU
: TELEPHONE: (202)672-5300
: TELEFAX: (202)672-5399
: TELEX: 904136
: INFORMATION FOR SEQ ID NO: 52:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4145 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Other nucleic acid;
: DESCRIPTION: Synthetic DNA oligonucleotide
: IMMEDIATE SOURCE:
: CLONE: pS2qpt-S4
: US-08-358-928-62

Query Match 5.0%; Score 68; DB 4; Length 4145;
Best Local Similarity 100.0%; Pred. No. 6.9e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

Oy 1299 gccaccgcggtgagctccagcttttgccttttagtgagggttaattgcgcgttgcc 1358
|||||
Db 2219 GCCACCGCGGTGAGCTCCAGCTTTTGTCCCTTTAGTAGAGGGTTAAITGCCGCTTGGC 2150

Oy 1359 gtaatacat 1366
|||||
Db 2159 GTAATCAT 2152

RESULT 18
US-08-651-472-63/c
: Sequence 63, Application US/08651472
: Patent No. 5103244
: GENERAL INFORMATION:
: APPLICANT: DORNER, Friedrich
: APPLICANT: SCHEIFLINGER, Friedrich
: APPLICANT: FALKNER, Falko Gunter
: APPLICANT: PFLEIDERER, Michael
: TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF CHIMERIC
: TITLE OF INVENTION: VIRUSES CONTAINING HUMAN IMMUNODEFICIENCY VIRUS TYPE 1
: NUMBER OF SEQUENCES: 95

```

```

: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W., Suite 500
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/651.472
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/914,738
: FILING DATE: 20-JUL-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/750,080
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/156/IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)672-5300
: TELEFAX: (202)672-5399
: TELEX: 904136
: INFORMATION FOR SEQ ID NO: 63:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4277 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Other nucleic acid;
: DESCRIPTION: Synthetic DNA oligonucleotide
: IMMEDIATE SOURCE:
: CLONE: pS2qpt-P2
: US-08-651-472-63

Query Match 5.0%; Score 68; DB 3; Length 4277;
Best Local Similarity 100.0%; Pred. No. 6.9e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

Oy 1299 gccaccgcggtgagctccagcttttgccttttagtgagggttaattgcgcgttgcc 1358
|||||
Db 2219 GCCACCGCGGTGAGCTCCAGCTTTTGTCCCTTTAGTAGAGGGTTAAITGCCGCTTGGC 2160

Oy 1359 gtaatacat 1366
|||||
Db 2159 GTAATCAT 2152

RESULT 19
US-08-358-928-63/c
: Sequence 63, Application US/08358928
: Patent No. 6265183
: GENERAL INFORMATION:
: APPLICANT: DORNER, Friedrich
: APPLICANT: SCHEIFLINGER, Friedrich
: APPLICANT: FALKNER, Falko Gunter
: APPLICANT: PFLEIDERER, Michael
: TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF CHIMERIC
: TITLE OF INVENTION: VIRUSES CONTAINING HUMAN IMMUNODEFICIENCY VIRUS TYPE 1
: NUMBER OF SEQUENCES: 95
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W., Suite 500
: CITY: Washington

```


STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/358,928
FILING DATE: 25-AUG-1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/914,738
FILING DATE: 20-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/750,080
FILING DATE: 25-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/165/IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 4277 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Other nucleic acid;
DESCRIPTION: Synthetic DNA oligonucleotide
IMMEDIATE SOURCE:
CLONE: pS29pt-P2
US-08-358-928-63

Query Match 5.0%; Score 68; DB 4; Length 4277;
Best Local Similarity 100.0%; Pred. No. 6.9e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1299 gccaccgcggtgagctccagcttttgcctttagtgaggttaattgcgcgcttgac 1358
|||||
Db 2219 gccaccgcggtgagctccagcttttgcctttagtgaggttaattgcgcgcttgac 1358
|||||

Oy 1359 gtaatcat 1366
|||||
Db 2159 gtaatcat 2152
|||||

RESULT 20
US-08-992-334-2
Sequence 2, Application US/08992334
Patent No. 5919678
GENERAL INFORMATION:
APPLICANT: Gruss, Alexandra
APPLICANT: Maguin, Emmanuelle
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
TITLE OF INVENTION: PLASMID
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESS: Christie Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992.334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/03034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5234 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-992-334-2

Query Match 5.0%; Score 68; DB 2; Length 5234;
Best Local Similarity 100.0%; Pred. No. 6.7e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1299 gccaccgcggtgagctccagcttttgcctttagtgaggttaattgcgcgcttgac 1358
|||||
Db 5000 gccaccgcggtgagctccagcttttgcctttagtgaggttaattgcgcgcttgac 5059
|||||

Oy 1359 gtaatcat 1366
|||||
Db 5060 gtaatcat 5067
|||||

RESULT 21
US-08-302-752-2
Sequence 2, Application US/08302752
Patent No. 6025190
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: THERMOSENSIBLE PLASMID
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5234 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

us-09-248-178-61.rni

Tue Apr 30 14:18:55 2002

MOLECULE TYPE: DNA (genomic)
US-08-302-752-2

Query Match 5.0%; Score 68; DB 3; Length 5234;

Best Local Similarity 100.0%; Pred. No. 6.7e-18;

Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1299 gccaccggtgagctccagcttttcttccttagtgagggttaattgcgcgttggc 1358

|||||

Db 5000 GCCACCGCGGTGGAGCTCCAGCTTTTGTTCCTTTAGTACAGGGTTAAATGGCGCTTGCC 5059

Oy 1359 gtaatcat 1366

|||||

Db 5060 GTAATCAT 5067

RESULT 22

US-08-446-935-1

Sequence 1, Application US/08446935

Patent No. 6187991

GENERAL INFORMATION:

APPLICANT: Soeller, Walter C.

APPLICANT: Carly, Maynard D.

APPLICANT: Kreutter, David K.

TITLE OF INVENTION: TRANSGENIC ANIMAL MODELS FOR TYPE II

TITLE OF INVENTION: DIABETES MELLITUS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pfizer Inc.

STREET: 235 East 42nd Street, 20th Floor

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10017-5755

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/446,935

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Shevka, Robert F.

REGISTRATION NUMBER: 31,304

REFERENCE/DOCKET NUMBER: PC8153

TELEPHONE: (212)573-1189

TELEFAX: (212)573-1939

TELEX: N/A

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 5356 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

US-08-446-935-1

Query Match 5.0%; Score 68; DB 4; Length 5356;

Best Local Similarity 100.0%; Pred. No. 6.7e-18;

Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1299 gccaccggtgagctccagcttttcttccttagtgagggttaattgcgcgttggc 1358

|||||

Db 3138 GCCACCGCGGTGGAGCTCCAGCTTTTGTTCCTTTAGTACAGGGTTAAATGGCGCTTGCC 197

Oy 1359 gtaatcat 1366

|||||

Db 3198 GTAATCAT 3205

RESULT 23

US-08-651-472-72/c

Sequence 72, Application US/08651472

Patent No. 6103244

GENERAL INFORMATION:

APPLICANT: DORNER, Friedrich

APPLICANT: SCHEIFLINGER, Friedrich

APPLICANT: FALKNER, Faiko Gunter

APPLICANT: PFELEIDERER, Michael

TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF CHIMERIC

TITLE OF INVENTION: VIRUSES CONTAINING HUMAN IMMUNODEFICIENCY VIRUS TYPE 1

TITLE OF INVENTION: (HIV-1) ANTIGENS

NUMBER OF SEQUENCES: 95

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/651,472

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/914,738

FILING DATE: 20-JUL-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/750,080

FILING DATE: 26-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472/166/IMMU

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 72:

SEQUENCE CHARACTERISTICS:

LENGTH: 5532 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Other nucleic acid;

DESCRIPTION: Synthetic DNA oligonucleotide

IMMEDIATE SOURCE:

CLONE: pAN2gpta-FIX

US-08-651-472-72

Query Match 5.0%; Score 68; DB 3; Length 5532;

Best Local Similarity 100.0%; Pred. No. 6.7e-18;

Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1299 gccaccggtgagctccagcttttcttccttagtgagggttaattgcgcgttggc 1358

|||||

Db 2219 GCCACCGCGGTGGAGCTCCAGCTTTTGTTCCTTTAGTACAGGGTTAAATGGCGCTTGCC 2160

Oy 1359 gtaatcat 1366

|||||

Db 2159 GTAATCAT 2152

RESULT 24

US-08-358-928-72/c

; Sequence 72, Application US/08358928
; Patent No. 6265183
; GENERAL INFORMATION:
; APPLICANT: DORNER, Friedrich
; APPLICANT: SCHEIFLINGER, Friedrich
; APPLICANT: FALKNER, Falko Gunter
; APPLICANT: PFLEIDERER, Michael
; TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF CHIMERIC
; TITLE OF INVENTION: VIRUSES CONTAINING HUMAN IMMUNODEFICIENCY VIRUS TYPE 1
; TITLE OF INVENTION: (HIV-1) ANTIGENS
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/358,928
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/914,738
; FILING DATE: 20-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/750,080
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/166/IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5532 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid;
; DESCRIPTION: Synthetic DNA oligonucleotide
; IMMEDIATE SOURCE:
; CLONE: pN29pta-FIX
; US-08-358-928-72

Query Match 5.0%; Score 68; DB 4; Length 5532;
Best Local Similarity 100.0%; Pred. No. 6.7e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1299 gccaccgcggtgagctccagcttttttccctttagtgagggttaattgcgcgttggc 1358
Db 2219 GCCACCGCGGTGGAGCTCCAGCTTTTGTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGC 2160
OY 1359 gtaatcat 1365
Db 2159 GTAATCAT 2152

RESULT 25
US-08-992-334-3
; Sequence 3, Application US/08992334
; Patent No. 5919678
; GENERAL INFORMATION:
; APPLICANT: Gruss, Alexandra

; APPLICANT: Maguin, Emmanuelle
; TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
; TITLE OF INVENTION: PLASMID
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christie Parker & Hale, LLP
; STREET: 350 West Colorado Boulevard, Suite 500
; CITY: Pasadena
; STATE: California
; COUNTRY: United States
; ZIP: 91105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/992,334
; FILING DATE: 17-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,752
; FILING DATE: 24-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00248
; FILING DATE: 12-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/03034
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Prout, D. Bruce
; REGISTRATION NUMBER: 20958
; REFERENCE/DOCKET NUMBER: C93:31779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 795-9900
; TELEFAX: (626) 577-8800
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6722 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; US-08-992-334-3

Query Match 5.0%; Score 68; DB 2; Length 6722;
Best Local Similarity 100.0%; Pred. No. 6.5e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1299 gccaccgcggtgagctccagcttttttccctttagtgagggttaattgcgcgttggc 1358
Db 6488 GCCACCGCGGTGGAGCTCCAGCTTTTGTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGC 6547
OY 1359 gtaatcat 1366
Db 6548 GTAATCAT 6555

RESULT 26
US-08-302-752-3
; Sequence 3, Application US/08302752
; Patent No. 6025190
; GENERAL INFORMATION:
; APPLICANT: THERMOSENSIBLE PLASMID
; TITLE OF INVENTION: THERMOSENSIBLE PLASMID
; NUMBER OF SEQUENCES: 3
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:

```

: APPLICATION NUMBER: US/08/302,752
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 9203034
: FILING DATE: 13-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO FR/93/00248
: FILING DATE: 12-MAR-1993
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6722 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-302-752-3

Query Match          5.0%  Score 68; DB 3; Length 6722;
Best Local Similarity 100.0%; Pred. No. 6.5e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1299 gccaccgcggtgagctccagctttttcccttttaagagatttaattgcgcgttggc 1358
      |||||||
Db 6488 gccaccgcggtgagctccagctttttcccttttaagagatttaattgcgcgttggc 1358

Qy 1359 gtaatcat 1366
      |||||||
Db 6548 gtaatcat 6555

RESULT 27
US-08-651-472-67/c
: Sequence 67, Application US/08651472
: Patent No. 6103244
: GENERAL INFORMATION:
: APPLICANT: DORNER, Friedrich
: APPLICANT: SCHEIFLINGER, Friedrich
: APPLICANT: FALKNER, Falko Gunter
: APPLICANT: PFLEIDERER, Michael
: TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF CHIMERIC
: TITLE OF INVENTION: VIRUSES CONTAINING HUMAN IMMUNODEFICIENCY VIRUS TYPE 1
: TITLE OF INVENTION: (HIV-1) ANTIGENS
: NUMBER OF SEQUENCES: 95
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W., Suite 500
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/651,472
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/914,738
: FILING DATE: 20-JUL-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/750,080
: FILING DATE: 26-AUG-1991
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/166/IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)672-5300
: TELEFAX: (202)672-5399
: INFORMATION FOR SEQ ID NO: 67:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6811 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Other nucleic acid;
: DESCRIPTION: Synthetic DNA oligonucleotide
: IMMEDIATE SOURCE:
: CLONE: DN2-qpta Prots
: US-08-651-472-67

TELEFAX: (202)672-5399
TELEX: 904136
: INFORMATION FOR SEQ ID NO: 67:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6811 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Other nucleic acid;
: DESCRIPTION: Synthetic DNA oligonucleotide
: IMMEDIATE SOURCE:
: CLONE: DN2-qpta Prots
: US-08-651-472-67

Query Match          5.0%  Score 68; DB 3; Length 6811;
Best Local Similarity 100.0%; Pred. No. 6.5e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1299 gccaccgcggtgagctccagctttttcccttttaagagatttaattgcgcgttggc 1358
      |||||||
Db 2219 gccaccgcggtgagctccagctttttcccttttaagagatttaattgcgcgttggc 2160

Qy 1359 gtaatcat 1366
      |||||||
Db 2159 gtaatcat 2152

RESULT 28
US-08-358-928-67/c
: Sequence 67, Application US/08358928
: Patent No. 5265183
: GENERAL INFORMATION:
: APPLICANT: DORNER, Friedrich
: APPLICANT: SCHEIFLINGER, Friedrich
: APPLICANT: FALKNER, Falko Gunter
: APPLICANT: PFLEIDERER, Michael
: TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF CHIMERIC
: TITLE OF INVENTION: VIRUSES CONTAINING HUMAN IMMUNODEFICIENCY VIRUS TYPE 1
: TITLE OF INVENTION: (HIV-1) ANTIGENS
: NUMBER OF SEQUENCES: 95
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W., Suite 500
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/358,928
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/914,738
: FILING DATE: 20-JUL-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/750,080
: FILING DATE: 26-AUG-1991
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/166/IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)672-5300
: TELEFAX: (202)672-5399
: INFORMATION FOR SEQ ID NO: 67:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6722 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-302-752-3
```

LENGTH: 6811 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Other nucleic acid:
DESCRIPTION: Synthetic DNA oligonucleotide
IMMEDIATE SOURCE:
CLONE: pN2-gpLa Prots
US-08-358-928-67

Query Match 5.0%; Score 68; DB 4; Length 6811;
Best Local Similarity 100.0%; Pred. No. 6.5e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1299 gccacgcggtgagctccagcttttcccttttagtgaggttaattgcgcgttggc 1358
|||||
Db 2219 gccaccgggtggagctccagcttttcccttttagtgaggttaattgcgcgttggc 2160
|||||

Oy 1359 gtaatcat 1366
|||||
Db 2159 GTAATCAT 2152

RESULT 29
US-08-651-472-69/c
Sequence 69, Application US/08651472
Patent No. 6103244
GENERAL INFORMATION:
APPLICANT: DORNER, Friedrich
APPLICANT: SCHEIFLINGER, Friedrich
APPLICANT: FALKNER, Falko Gunter
APPLICANT: FELEIDERER, Michael
TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF CHIMERIC
TITLE OF INVENTION: VIRUSES CONTAINING HUMAN IMMUNODEFICIENCY VIRUS TYPE 1
TITLE OF INVENTION: (HIV-1) ANTIGENS
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/651,472
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/914,738
FILING DATE: 20-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/750,080
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/166/IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136

INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 6926 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: Other nucleic acid:
DESCRIPTION: Synthetic DNA oligonucleotide
IMMEDIATE SOURCE:
CLONE: pp2-gp160MN
US-08-651-472-69

Query Match 5.0%; Score 68; DB 3; Length 6926;
Best Local Similarity 100.0%; Pred. No. 6.5e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1299 gccacgcggtgagctccagcttttcccttttagtgaggttaattgcgcgttggc 1358
|||||
Db 2219 gccaccgggtggagctccagcttttcccttttagtgaggttaattgcgcgttggc 2160
|||||

Oy 1359 gtaatcat 1366
|||||
Db 2159 GTAATCAT 2152

RESULT 30
US-08-358-928-69/c
Sequence 69, Application US/08358928
Patent No. 6265183
GENERAL INFORMATION:
APPLICANT: DORNER, Friedrich
APPLICANT: SCHEIFLINGER, Friedrich
APPLICANT: FALKNER, Falko Gunter
APPLICANT: FELEIDERER, Michael
TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF CHIMERIC
TITLE OF INVENTION: VIRUSES CONTAINING HUMAN IMMUNODEFICIENCY VIRUS TYPE 1
TITLE OF INVENTION: (HIV-1) ANTIGENS
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/358,928
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/914,738
FILING DATE: 20-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/750,080
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/166/IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136

INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 6926 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Other nucleic acid:
DESCRIPTION: Synthetic DNA oligonucleotide
IMMEDIATE SOURCE:
CLONE: pp2-gp160MN

; Sequence 135, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgon
; APPLICANT: Knechtle, Philipp
; APPLICANT: Reibschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHEYA GOSSYPIL
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 6239264artis Corporation
; STREET: 3054 Cornwalis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 135:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 713 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1051RP
; US-08-998-416-135

Query Match 4.5%: Score 62; DB 4; Length 713;
Best Local Similarity 100.0%; Pred. No. 1.8e-15;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ccagtgagcgcgcgtaaacgactcactataggcgcaattgggtaccgggccccccctcg 60
Db 246 CCAGTGAGCGCGCGTAATACGACTACACTATAGGGCGGAATTGGTACCGGGCCCCCTCG 187
Qy 61 ag 62
Db 186 AG 185

RESULT 34
US-08-452-267-3/C
; Sequence 3, Application US/08452267
; Patent No. 5801027
; GENERAL INFORMATION:
; APPLICANT: Bennett, Malcolm
; APPLICANT: May, Sean
; APPLICANT: Ramsay, Nicola
; TITLE OF INVENTION: Control of Genes in Transgenic Plants

; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Barnes & Thornburg
; STREET: 11 South Meridian
; CITY: Indianapolis
; STATE: IN
; COUNTRY: USA
; ZIP: 46204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,267
; FILING DATE: 26-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Breen, John P.
; REGISTRATION NUMBER: 38,833
; REFERENCE/DOCKET NUMBER: 6653-25744
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 231-7745
; TELEFAX: (317) 231-7433
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5534 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: PUMIGIT
; US-08-452-267-3
Query Match 4.5%: Score 62; DB 1; Length 5534;
Best Local Similarity 100.0%; Pred. No. 1.4e-15;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ccagtgagcgcgcgtaaacgactcactataggcgcaattgggtaccgggccccccctcg 60
Db 2717 CCAGTGAGCGCGCGTAATACGACTACACTATAGGGCGGAATTGGTACCGGGCCCCCTCG 2658
Qy 61 ag 62
Db 2657 AG 2656
RESULT 35
US-09-123-644-3/C
; Sequence 3, Application US/09123644
; Patent No. 5127606
; GENERAL INFORMATION:
; APPLICANT: Bennett, Malcolm
; APPLICANT: May, Sean
; APPLICANT: Ramsay, Nicola
; TITLE OF INVENTION: Method of Using Transactivation Proteins to
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Barnes & Thornburg
; STREET: 11 South Meridian
; CITY: Indianapolis
; STATE: IN
; COUNTRY: USA
; ZIP: 46204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/123,644
;; FILING DATE: 28-JUL-1998
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Breen, John P.
;; REGISTRATION NUMBER: 38,833
;; REFERENCE/DOCKET NUMBER: 5653-60788
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (317) 231-7745
;; TELEFAX: (317) 231-7433
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5534 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: circular
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; IMMEDIATE SOURCE:
;; CLONE: pUMIGIT
US-09-123-644-3

Query Match 4.5% Score 62: DB 3: Length 5534;
Best Local Similarity 100.0%; Pred. No. 1.4e-15;
Matches 62: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccagtgagcgcgtaatacagactcactataggcggaattgggtacgggccccccctcg 60
DB 2717 CCAGTGAGCGCGTAATAACGACTACTATAGGCGGAATTGGGTACGGGCGGCCCTCG 2658
OY 61 ag 62
DB 2657 AG 2656

RESULT 36
US-09-128-314-4/c
; Sequence 4, Application US/09128314
; Patent No. 6183121
; GENERAL INFORMATION:
; APPLICANT: Kim, Joseph L
; APPLICANT: Morgenstern, Kurt A
; APPLICANT: Caron, Paul R
; APPLICANT: Lin, Chao
; APPLICANT: Vertex Pharmaceuticals Inc.
; TITLE OF INVENTION: CRYSTAL STRUCTURE OF THE HCV NS3 HELICASE DOMAIN
; FILE REFERENCE: Sequence listing for VPI/97-101
; Patent No. 6183121
; CURRENT APPLICATION NUMBER: US/09/128,314
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: 60/055,772
; EARLIER FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 7659
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Genetically
; OTHER INFORMATION: engineered plasmid containing helicase domain of
; OTHER INFORMATION: HCV NS3
US-09-128-314-4

Query Match 4.5% Score 62: DB 4: Length 7659;
Best Local Similarity 100.0%; Pred. No. 1.3e-15;
Matches 62: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ccagtgagcgcgtaatacagactcactataggcggaattgggtacgggccccccctcg 60

DB 1900 CCAGTGAGCGCGTAATAACGACTACTATAGGCGGAATTGGGTACGGGCCCCCTCG 1841
OY 61 ag 62
DB 1840 AG 1839
RESULT 37
US-09-128-314-3/c
; Sequence 3, Application US/09128314
; Patent No. 6183121
; GENERAL INFORMATION:
; APPLICANT: Kim, Joseph L
; APPLICANT: Morgenstern, Kurt A
; APPLICANT: Caron, Paul R
; APPLICANT: Lin, Chao
; APPLICANT: Vertex Pharmaceuticals Inc.
; TITLE OF INVENTION: CRYSTAL STRUCTURE OF THE HCV NS3 HELICASE DOMAIN
; FILE REFERENCE: Sequence listing for VPI/97-101
; Patent No. 6183121
; CURRENT APPLICATION NUMBER: US/09/128,314
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: 60/055,772
; EARLIER FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 8157
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Genetically
; OTHER INFORMATION: engineered plasmid containing full-length HCV NS3
; OTHER INFORMATION: coding sequence
US-09-128-314-3

Query Match 4.5% Score 62: DB 4: Length 8157;
Best Local Similarity 100.0%; Pred. No. 1.3e-15;
Matches 62: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccagtgagcgcgtaatacagactcactataggcggaattgggtacgggccccccctcg 60
DB 2398 CCAGTGAGCGCGTAATAACGACTACTATAGGCGGAATTGGGTACGGGCCCCCTCG 2339
OY 61 ag 62
DB 2338 AG 2337

RESULT 38
US-09-096-947-2/c
; Sequence 2, Application US/08096947
; Patent No. 5506137
; GENERAL INFORMATION:
; APPLICANT: Eric J. Mathur
; APPLICANT: Edward J. Marsh
; APPLICANT: Warren E. Schoettlin
; TITLE OF INVENTION: Purified thermostable
; TITLE OF INVENTION: Pyrococcus furiosus DNA
; TITLE OF INVENTION: Ligase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Limbach & Limbach
; STREET: 2001 Ferry Building
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3+ inch, 1.4 Mb storage
; COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,947
FILING DATE: 19930722
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/919,140
FILING DATE: July 23, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Albert P. Halluin
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: STRG 20081 USA
TELEPHONE: (415) 433-4150
TELEFAX: (415) 433-8715
TELEX: 278356
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2363 base pairs
TYPE: nucleic acid
STRAINEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Pyrococcus Furiosus
INDIVIDUAL ISOLATE: DSM #3638
CELL TYPE: unicellular organism
IMMEDIATE SOURCE:
LIBRARY: Pyrococcus Furiosus Genomic DNA
CLONE: pEM1
US-08-096-947-2

Query Match 4.5% Score 61: DB 1: Length 2363;
Best Local Similarity 100.0%; Pred. No. 3.7e-15;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1299 gccaccgcggtgagctccagctttgtcccttagtgagggttaattgcgcgttggc 1358
|||||
DB 141 GCCACCGCGGTGAGCTCCAGCTTTGTTCCCTTAGTGAGGGTTAATTCGCCGCTTGGC 82

QY 1359 g 1359
DB 81 G 81

RESULT 39
US-07-919-140B-2/c
Sequence 2, Application US/07919140B
Patent No. 5700672
GENERAL INFORMATION:
APPLICANT: Eric J. Mathur
APPLICANT: Edward J. Marsh
APPLICANT: Warren E. Schoettlin
TITLE OF INVENTION: Purified Thermostable Pyrococcus
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 2730 Sand Hill Road
CITY: Menlo Park
STATE: California
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/919,140B

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,947
FILING DATE: 19930722
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/919,140
FILING DATE: July 23, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Albert P. Halluin
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: STRG 20081 USA
TELEPHONE: (415) 433-4150
TELEFAX: (415) 433-8715
TELEX: 278356
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2363 base pairs
TYPE: nucleic acid
STRAINEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Pyrococcus Furiosus
INDIVIDUAL ISOLATE: DSM #3638
CELL TYPE: unicellular organism
IMMEDIATE SOURCE:
LIBRARY: Pyrococcus Furiosus Genomic DNA
CLONE: pEM1
US-08-096-947-2

Query Match 4.5% Score 61: DB 1: Length 2363;
Best Local Similarity 100.0%; Pred. No. 3.7e-15;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1299 gccaccgcggtgagctccagctttgtcccttagtgagggttaattgcgcgttggc 1358
|||||
DB 141 GCCACCGCGGTGAGCTCCAGCTTTGTTCCCTTAGTGAGGGTTAATTCGCCGCTTGGC 82

QY 1359 g 1359
DB 81 G 81

RESULT 39
US-07-919-140B-2/c
Sequence 2, Application US/07919140B
Patent No. 5700672
GENERAL INFORMATION:
APPLICANT: Eric J. Mathur
APPLICANT: Edward J. Marsh
APPLICANT: Warren E. Schoettlin
TITLE OF INVENTION: Purified Thermostable Pyrococcus
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 2730 Sand Hill Road
CITY: Menlo Park
STATE: California
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/919,140B

TELEFAX: (415) 433-8716
TELEX: 278356
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2363 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Pyrococcus Furiosus
INDIVIDUAL ISOLATE: DSM #3638
CELL TYPE: unicellular organism
IMMEDIATE SOURCE:
LIBRARY: Pyrococcus Furiosus Genomic DNA
CLONE: pEM1
PCT-US93-06939-2

Query Match 4.5% Score 61: DB 5: Length 2363:
Best Local Similarity 100.0%: Pred. No. 3.7e-15:
Matches 61: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Oy 1299 gccaccgcggtgagctccagcttttgcctttagtgagggttaattgcgcgttggc 1358
Db 141 gccaccgcggtgagctccagcttttgcctttagtgagggttaattgcgcgttggc 82

Oy 1359 g 1359
Db 81 G 81

RESULT 42
US-08-799-569-1
Sequence 1, Application US/08799569
Patent No. 6133244
GENERAL INFORMATION:
APPLICANT: Michel, Marie-Louise
TITLE OF INVENTION: Nucleotide Vector, Composition
TITLE OF INVENTION: Containing Such Vector, and Vaccine for Immunization
TITLE OF INVENTION: Against Hepatitis
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/799,569
FILING DATE: 12-FEB-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/706,337
FILING DATE: 30-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/633,821
FILING DATE: 22-APR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 94/00483
FILING DATE: 27-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J
REGISTRATION NUMBER: 25,146

REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 8142-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-854-3560
TELEFAX: 415-854-3594
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2363 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Pyrococcus Furiosus
INDIVIDUAL ISOLATE: DSM #3638
CELL TYPE: unicellular organism
IMMEDIATE SOURCE:
LIBRARY: PTO.MBPyrococcus Furiosus Genomic DNA
CLONE: pEM1
US-08-916-232-2

Query Match 4.5% Score 61: DB 4: Length 2363:
Best Local Similarity 100.0%: Pred. No. 3.7e-15:
Matches 61: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Oy 1299 gccaccgcggtgagctccagcttttgcctttagtgagggttaattgcgcgttggc 1358
Db 141 gccaccgcggtgagctccagcttttgcctttagtgagggttaattgcgcgttggc 82

Oy 1359 g 1359
Db 81 G 81

RESULT 41
PCT-US93-06939-2/C
Sequence 2, Application PC/TUS9306939
GENERAL INFORMATION:
APPLICANT: Eric J. Mathur
APPLICANT: Edward J. Marsh
APPLICANT: Warren E. Schoettlin
TITLE OF INVENTION: Purified Thermostable
TITLE OF INVENTION: Pyrococcus Furiosus DNA
TITLE OF INVENTION: Ligase
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Limbach & Limbach
STREET: 2001 Ferry Building
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3+ inch, 1.4 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/06939
FILING DATE: 19930722
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/919,140
FILING DATE: July 23, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Albert P. Halluin
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: STRG 20081 USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 433-4150

REFERENCE/DOCKET NUMBER: 03495.0128-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5618 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-799-569-1

Query Match 4.0%; Score 55; DB 3; Length 5618;
Best Local Similarity 100.0%; Pred. No. 5,8e-13;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1299 gccaccgggtgagctccagcttttattcccttttaagtgaggttaattgcgcac 1353
|||||
Db 2803 gccaccgggtgagctccagcttttattcccttttaagtgaggttaattgcgcac 2857

RESULT 43

US-08-703-809-9/c
Sequence 9, Application US/08703809
Patent No. 5716808
GENERAL INFORMATION:
APPLICANT: Raymond, Christopher K.
TITLE OF INVENTION: GENETIC ENGINEERING OF PICHIA METHANOLICA
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: ZymoGenetics, Inc.
STREET: 1201 Eastlake Avenue East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E
REGISTRATION NUMBER: 31-648
REFERENCE/DOCKET NUMBER: 96-18
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6673
TELEFAX: 206-442-6678
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 329 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

Query Match 3.7%; Score 51; DB 1; Length 329;
Best Local Similarity 100.0%; Pred. No. 3.4e-11;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1316 ccagcttttgccttttagtgagggtaattgcgcgttgcgtaatacat 1366
|||||
Db 55 CCAGCTTTTGTCCCTTTAGTGAGGGTTAATTCGCGCTTGCGGTAAATCAT 5

RESULT 44

US-08-703-808-9/c
Sequence 9, Application US/08703808
Patent No. 5736383
GENERAL INFORMATION:
APPLICANT: Raymond, Christopher K.
TITLE OF INVENTION: PREPARATION OF PICHIA METHANOLICA AUXOTROPHIC
MUTANTS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: ZymoGenetics, Inc.
STREET: 1201 Eastlake Avenue East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E
REGISTRATION NUMBER: 31-648
REFERENCE/DOCKET NUMBER: 96-17
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6673
TELEFAX: 206-442-6678
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 329 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

Query Match 3.7%; Score 51; DB 1; Length 329;
Best Local Similarity 100.0%; Pred. No. 3.4e-11;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1316 ccagcttttgccttttagtgagggtaattgcgcgttgcgtaatacat 1366
|||||
Db 55 CCAGCTTTTGTCCCTTTAGTGAGGGTTAATTCGCGCTTGCGGTAAATCAT 5

RESULT 45

US-08-703-807-9/c
Sequence 9, Application US/08703807
Patent No. 5955349
GENERAL INFORMATION:
APPLICANT: Raymond, Christopher K.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PRODUCING
HETEROLOGOUS POLYPEPTIDES IN PICHIA METHANOLICA
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: ZymoGenetics, Inc.
STREET: 1201 Eastlake Avenue East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE:

```
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31-548
; REFERENCE/DOCKET NUMBER: 96-16
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6573
; TELEFAX: 206-442-6578
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 329 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-703-807-9

Query Match      3.7%  Score 51; DB 2; Length 329:
Best Local Similarity 100.0%; Pred. No. 3.4e-11;
Matches 51: Conservative 0; Mismatches 0; Indels 0; Gaps 0:

Oy 1316 ccagctttgtccctttagtgagggttaattgcgcgttgacgtatcat 1366
    |||||||||||||||||||||||||||||||||||||||||||||||
Db 55 CCAGCTTTGTCCCTTTAGTGAGGGTTAATTGCCGCGCTTGGCGTAATCAT 5

Search completed: April 30, 2002, 10:38:12
Job time: 9961 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic : nucleic search, using sw model

Run on: April 29, 2002, 17:16:47 : Search time 5783.7 Seconds
(without alignments)
2721.794 Million cell updates/sec

Title: US-09-248-178-61
Perfect score: 1368
Sequence: 1 ccagtgagcgcgtaatac.....cgcccttgccgaatcatnn 1368

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207.seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_gss:*
 - 13: em_gss_hum:*
 - 14: em_gss_inv:*
 - 15: em_gss_pln:*
 - 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	551.8	40.3	556	9	AI927004 w068g06.x
2	516.4	37.7	576	9	AW374095 PM4-BT054
3	499.5	36.5	550	9	AW374082 PM4-BT054
4	393.4	28.8	905	10	BE872674 501451222
5	388.8	28.4	409	9	AI085849 0x47f12.s
6	352	25.7	364	9	AI867547 wb78c04.x
7	234.4	17.1	746	12	AG030921 Pan trogl
8	126.4	9.2	661	10	BM391501 UI-R-DY0-
9	122	8.9	636	10	BM384329 UI-R-DY0-
10	119	8.7	745	11	AK003832 Mus muscu
11	110.8	8.1	369	9	AW787180 120924 MA
12	99.4	7.3	733	10	BM391251 UI-R-DY0-
13	92.4	6.8	535	9	AW177437 AU177437
14	92	6.7	658	12	AG054901 Pan trogl
15	88.9	6.5	677	9	AK720437 C01755-F
16	87.2	6.4	471	9	AW660207 EST1372278
17	85.2	6.2	311	12	A2049564 GSSBfru013

C 18	83	6.1	653	9	AU177453
C 19	82.2	6.0	627	10	BE414300
C 20	82.2	6.0	800	10	BE421927
C 21	81.6	6.0	253	10	BI784504
C 22	81.6	6.0	358	9	AU069050
C 23	81.6	6.0	363	9	AU069112
C 24	81.6	6.0	372	10	D80022
C 25	81.6	6.0	394	9	AU069190
C 26	81.2	5.9	385	9	AW951430
C 27	81.2	5.9	450	10	D59275
C 28	81.2	5.9	945	12	CNS04D0K
C 29	80.6	5.9	325	9	AU069730
C 30	80.6	5.9	467	9	AV746034
C 31	80.6	5.9	640	10	BE414138
C 32	80	5.8	438	10	BI945663
C 33	80	5.8	520	10	D50995
C 34	79.8	5.8	233	12	A0012365
C 35	79.6	5.8	311	9	AU068988
C 36	79.6	5.8	403	9	AW951437
C 37	79.6	5.8	421	9	AW951452
C 38	79	5.8	327	9	AU069088
C 39	79	5.8	329	10	BI118265
C 40	79	5.8	332	9	AU069089
C 41	79	5.8	476	10	BG662546
C 42	79	5.8	480	9	AW687437
C 43	79	5.8	493	9	AW954068
C 44	79	5.8	711	10	BE414359
C 45	79	5.8	732	10	BE421895

ALIGNMENTS

RESULT 1
LOCUS AI927004 556 bp mRNA linear EST 08-NAR-2000
DEFINITION w068g06.x1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:2460538 3', mRNA sequence.

ACCESSION AI927004
VERSION AI927004
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 556)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Sequencing by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 989 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 454.
Location/Qualifiers
1...556
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2460538"
/clone_lib="NCI_CGAP_Pr22"
/sex="male"
/tissue.type="normal prostate"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)

FEATURES
source

[illegible]

```
Db 704 ACAATTGTTGGGATTAAATGGTCGAGATGTGCGCCGCTGTCCTCGTTCAGTTCATTAT 645
Oy 705 tccgggtcgaaacttccaaaaaagggtt--ggcatgaagaaaaaataatcmaatca 761
Db 644 TCGGGCTGGGAAGCGTCGCGATAAAGGGGTAGGCTTGAGAGTGTAGTAAATCGAATCA 585
Oy 762 gaacctcttcaggaggttggkctctgatatggcagacargatatacaagctccaccagaga 821
Db 584 GAACCTCTTCAGGGGTTT--GTTCGTGATATGGCAGACAAAGTACAAAGTCCACCAGGAGA 527
Oy 822 tggagcaattcaaaataagggtaattggctgacaaaggtattattgccaagatgagacaga 881
Db 526 TGGAGCAATTTCAGATATAGGGTAATGGGCTGACTAGGTATTATGSCCAGCATGGGACAGA 467
Oy 882 atgagcaacagcgtcaaaagtttttggattatatagcsccttagagctctctgatatagga 941
Db 466 ATGAGCAGAGGCTGAAAGGTTGTGTGGATTATATAGCACCTAGAGTCTCTCAITGATGGGA 407
Oy 942 attttgttagtcaaacatcacctaaacttccaagggaaaaacttctcaggtagcctaadc 1001
Db 406 ATTITTTGTTAGTCAACATACGCTAACTTCCAAGGGGAAATCTTTCAGGTAGCCTAAGC 347
Oy 1002 ttgctttctagagtgatgagtgctgctactgtctgctgatttttgaacaaactgggtt 1061
Db 346 TTGCTTTCTAGAGTCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 288
Oy 1062 tgtacaagtgagaaagactagagagaaagatttttagctctgtatttagcagagccattttat 1121
Db 287 TGTAC--AGTGAGAAAGACTAGAGAGAAAGATTTTAGTCTGTTTAGCAGAGGCCAATA 229
Oy 1122 ctgctgacatgagatcaatattctctgctccctatccccaggaagggaacaaatcccaa 1181
Db 228 CTGCGTGACATGATCAATATTTCTGATCCCTATATCCCGAGGAAGGCAAA--TCCCA 171
Oy 1182 agaatgtgttagcaaaatggctgctgctatcatatttctctgagcattgtctttg 1238
Db 170 AGAAATGTGTAGC--AAATGGCTGATGCTATCAATATGCTATGCTATGCTATGCTATG 115

RESULT 5
AI085849 409 bp mRNA linear EST 28-AUG-1998
LOCUS ox47f12.s1 Soares_total_fetus_Nb2HF8_9* Homo sapiens cDNA clone
DEFINITION IMAGE:1659503 3', mRNA sequence.
ACCESSION AI085849
VERSION AI085849
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 409)
JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Insert Length: 1869 Std Error: 0.00
Seq Primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 398.
Location/Qualifiers
1..409
/db_xref="taxon:9606"
/clone="IMAGE:1659503"
/dev_tag="8-9 weeks"
/lab_host="DH10B"
/note="vector: p7T3D-Pac (Pharmacia) with a modified
polylinker: Site_1: Not 1; Site_2: Eco RI; 1st strand cDNA
```

was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dI) primer [5'-TGTTACCAATCTGAAGTGGAGCGCCCTTAATTTTCTTTTCTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. *

BASE COUNT 158 a 58 c 59 g 134 t

ORIGIN

Query Match 28.4%; Score 388.8; DB 9; Length 409;

Best Local Similarity 99.5%; Pred. No. 1.6e-55;

Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 82 ttttttattgacagaattcaggctttattattgagcaatgaaaaacagctaaaaacttaatt 141

Db 1 TTTTATTGATCAGAAATTCAGGCTTTATTATTGAGCAATGAAACAGCTTAAACCTTAAT 60

Oy 142 tccaagcatgtagtttaaagtttgcaagtgaggatattttcacaaaaacacacattcaatg 201

Db 51 TCCAAGCATGTAGTTAAAGTTTGCAGAGTGGGATATTGTCACAAAACACATTCGAATG 120

Oy 202 ttttaacactatttttgaagaacaaaatatatttaaaattgttcttcttaaaagcc 261

Db 121 TTTAAACACTATTTTATTGGAAGAACAAAATATATTAAAAATTGTTCTTCTTAAAGGCC 180

Oy 262 cattctctcgaagctaaactttgttaattgtatttaagcaatgaagttattttgtaca 321

Db 181 CATTTCCCTCCNAGCTTAAAGTTTGTAAATTTGATATTAGCAATGAAGTTATTTTGTACA 240

Oy 322 atctagttaaacaaacagagaatagcactaggcagaataaaaaattgcacagactatcaaa 381

Db 241 ATCTAGTTAAACAAGCAGAAATAGCACTAGGCAGAAATAAAAAATTGCACAGCGTATGCCAA 300

Oy 382 ttttccaagatagcatcttttaaaattcagtttccagcttccaaagattggttgcctataa 441

Db 301 TTTTCCAAGATAGCATCTTTTAAATTCAGTATTCAGCTTCCANAGATTGTTGGGCATAA 360

Oy 442 tagacttaacataataatgatggctaaaaaaa 473

Db 361 TAGACTTAAACATATAATGATGGCTAAAAAAA 392

RESULT 6

AI957547

LOCUS wB78c04.x1 NC1-CGAP_Pr28 Homo sapiens cDNA clone

DEFINITION mRNA sequence.

ACCESSION AI957547

VERSION AI957547.1 GI:5540656

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE 1 (bases 1 to 364)

JOURNAL NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov

Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Library Arrayed by: Greg Lennon, Ph.D.

Cloning Distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 903 Std Error: 0.00


```

Seq primer: -40UP from Gibco
High quality sequence stop: 362.
Location/Qualifiers
1. 362
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2311782"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: Prostate; Vector: p73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonelids
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Ronaldo."
BASE COUNT 134 a 55 c 52 g 122 t
ORIGIN
Query Match 25.7%; Score 352; DB 9; Length 362;
Best Local Similarity 99.7%; Pred. No. 2.4e-49;
Matches 363; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
Oy 85 ttattgacgaattcaggctttattattgagcaatgaaacagctaaaccttaattcc 144
Db 1 tttattgacgaattcaggctttattattgagcaatgaaacagctaaaccttaattcc 50
Oy 145 aagcatgttagttaaagttgcaagtggtgatttgcacaaacacattcaatgttt 204
Db 61 AAGCATGTGTAGTTAAAGTTTGCAAGTGGGATATTGTTCAACAAACACATTCAATGTT 120
Oy 205 aacactatttattgaagaacaaatatatttaaattgttctcctcaaaagcccat 264
Db 121 AAMACATATTATTGAAGAACAAATATATTAAATTTGTTCTCTAAAGCCCAT 180
Oy 265 tccccccaagctcaactttgtaattgatatgaagcaatgaagtattttgtcaatc 324
Db 181 TTCCCTCCAGTCTAACTTTGTAATTTGATATTGAAGCAATGAAGTTATTTGTACAAATC 240
Oy 325 tagttaaacagacagatagcactaagcagaataaaattcacagacgtatgcaattt 384
Db 241 TAGTTAAACAGCAGAAATAGCAGTAGCAGCAATAAAATTTGCACAGACGATGCANATTT 300
Oy 385 tccaagatagcattctttaaattcagtttgcagcttcccaaga-ttgggtgcccataa 443
Db 301 TCCAGATACCATTCITTTAAATTCAGTTTTCAGCTTCCAAAGATTGTTGCCATAATA 360
Oy 444 gact 447
Db 361 GACT 364

RESULT 7
AG030921 AG030921 746 bp DNA linear GSS 01-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-003H07.F, genomic survey sequence.
DEFINITION AG030921
ACCESSION AG030921
VERSION AG030921.1 GI:15557794
KEYWORDS GSS; GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC library clone:PTB-003H07.F.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
1 (sites)
Fujiyama.A., Hattori.M., Toyoda.A., Taylor.T.D., Yada.T.,
Tokoi.Y., Watanabe.H. and Sakaki.Y.
BAC end sequences of library PTB

REFERENCE
AUTHORS
TITLE

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 746)
AUTHORS Fujiyama.A., Hattori.M., Toyoda.A., Taylor.T.D., Yada.T.,
Tokoi.Y., Watanabe.H. and Sakaki.Y.
DIRECT SUBMISSION
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbesc@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1. 746
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-003H07.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 213 a 173 c 135 g 224 t 1 others
ORIGIN
Query Match 17.1%; Score 234.4; DB 12; Length 746;
Best Local Similarity 97.5%; Pred. No. 7.4e-30;
Matches 238; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Oy 82 ttttttattgacgaattcaggctttattattgagcaatgaaacagctaaaccttaatt 141
Db 503 TTTTCTATTGATGACGAATTCAGGCTTTATTATTGAGCAATGAAACAGCTAAACCTTAT 562
Oy 142 tccaagcatgttagttaaagttgcaagtggtgatttgcacaaacacattcaatg 201
Db 563 TCCAGCATGTGTAGTTAAAGTTTGCAAGTGGGATATTGTTCAACAAACACATTCAATG 622
Oy 202 ttttaaacactatttatttgaagaacaaatatattttaaattgttctctcaaaagcc 261
Db 623 TTAAACACTGTGTTATTATTGAAGAACAAATATATTTAAATTTGTTCTCTAAAGCC 682
Oy 262 catctccctcctcaagctaaacctttgttaatttgatattgaagcaatgaagtattttgtaca 321
Db 583 CATTTCCCTCCAGCTCAAACTTTGTAATTTGATATTGAAGCAATGAAGCTATTATTACA 742
Oy 322 atct 325
Db 743 GTCT 745

RESULT 8
BM391501 BM391501 661 bp mRNA linear EST 17-JAN-2002
LOCUS UI-R-DY0-ckr-e-06-0-UI.sl UI-R-DY0 Rattus norvegicus CDNA clone
DEFINITION UI-R-DY0-ckr-e-06-0-UI 3', mRNA sequence.
ACCESSION BM391501
VERSION BM391501
KEYWORDS EST.
SOURCE Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 661)
Bonaldo.M.F., Lennon.G. and Soares.M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery

```


Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 413-491, >R1_MM=SINE/Alu 436-491, >PB1D10SINE/Alu 511-570, >ID3=SINE/1D
Seq primer: M13 Forward
polyA=Yes.

FEA*JPEs	source	Location/Qualifiers			
		1. .733			
		/organism="Rattus norvegicus"			
		/strain="Sprague-Dawley"			
		/db_xref="taxon:10116"			
		/clone="UI-R-DY0-ckg-k-16-0-UI"			
		/clone_lib="UI-R-DY0"			
		/dev_stage="adult"			
		/lab_host="DHI0B (Life Technologies)"			
		/note="vector: pT773D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; UI-R-DY0 is a non-normalized rat cartilage library (RC) constructed in pT317 PAC vector according to the procedure described by Bonaldo, Lennon & Soares (Genome Research Genome 5: 731-806, 1996). The oligonucleotide used to prime first strand synthesis contained the sequence tad CTAATGGACG between the Not I cloning site and d18 stretch. The Rat cartilage tissue was provided by Dr Jeff Stevens at the University of Iowa.			
		TAG_L1B=UI-R-DY0			
		TAG_TISSUE=cartilage			
		TAG_SEQ=CTAATGGACG"			
214 a	149 c	137 g	232 t	i	others
BASE COUNT					
ORIGIN					

```

Query Match      7.3%: Score 99.4; DB 10; Length 733:
Best Local Similarity 60.1%: Pred. No. 2.6e-07:
Matches 224; Conservative 0; Mismatches 136; Indels 13; Gaps 3:

Qy 211 tatttattgaagacaaatatatttaaattgttgctctcaaaagccattccct 270
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10 TTTTATTGAGAGACAGATATATTCAGACITGTTATTCCACAAAACCAITTT-CCT 68

Qy 271 ccaagtcctaaattgtaattgataattagaacatgaagta-ttttgtacaattctagt 329
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 69 CAAGCCCTAAATTTATGCTGTGTTTAAGCAATGAAGTTATTTTATAAAACCTAGTII 128

Qy 330 aaacagacagataagcactaggcagaataaaaaattgcacagacgtatgcgaatttccaa 389
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 129 AAGCAATTTAATGGNAATAGCATCGTTAGATAAAAAATTCACACATGCGTITACTTCTCA 188

Qy 390 gatagcattctttaaattcagtttttcagcttccaaagatttggttcccaataataga-ctta 449
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 189 AATAGTATTATTITGATGCCAACITTTTCAGCTTCTAGTACAAAATTCGCTTCTAATAGACTGA 248

Qy 450 aacataataatgatgctcaaaaaataagatatacgaataatgtaaaaaaggaatgtaagtt 509
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 249 AATGCATAATGATAGTTTGAAGAAATATATATGTAAAGAAAGTAAA-----AATA 297

Qy 510 ccactctcaactctcaaaaggtgagatgaggaatgcttaagcagcaaaataaattgaggttc 569
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 298 CATTGCTGTCATATATAAAGGAGAGACCGTAAAGGATACACACGTGACAGAAAGGGAATTT 357

Qy 570 ttttttctgattt 582
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 358 TTGTTTTTGCTTT 370
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

DEFINITION      AUI77437 OLHNI cell line cDNA library insert; ESI 21-MAR-2001
LOCUS           AUI77437/c
FEATURES         535 bp mRNA
DEFINITION      clone OL601.9lb similar to pirlf68534; ribosomal protein L34 -
                 human, mRNA sequence.
ACCESSION       AUI77437.1 GI:13426273
VERSION         AUI77437.1

```

KEYWORDS	REFERENCE
SUPCE	AUTHORS
ORGANISM	TITLE
	JOURNAL
	COMMENT

FEATURES
SOURCE

DATE	TIME	BASE COUNT	CRIGIN
11/11/54	11:00	100	100
11/11/54	11:05	100	100
11/11/54	11:10	100	100
11/11/54	11:15	100	100
11/11/54	11:20	100	100
11/11/54	11:25	100	100
11/11/54	11:30	100	100
11/11/54	11:35	100	100
11/11/54	11:40	100	100
11/11/54	11:45	100	100
11/11/54	11:50	100	100
11/11/54	11:55	100	100
11/11/54	12:00	100	100
11/11/54	12:05	100	100
11/11/54	12:10	100	100
11/11/54	12:15	100	100
11/11/54	12:20	100	100
11/11/54	12:25	100	100
11/11/54	12:30	100	100
11/11/54	12:35	100	100
11/11/54	12:40	100	100
11/11/54	12:45	100	100
11/11/54	12:50	100	100
11/11/54	12:55	100	100
11/11/54	13:00	100	100
11/11/54	13:05	100	100
11/11/54	13:10	100	100
11/11/54	13:15	100	100
11/11/54	13:20	100	100
11/11/54	13:25	100	100
11/11/54	13:30	100	100
11/11/54	13:35	100	100
11/11/54	13:40	100	100
11/11/54	13:45	100	100
11/11/54	13:50	100	100
11/11/54	13:55	100	100
11/11/54	14:00	100	100
11/11/54	14:05	100	100
11/11/54	14:10	100	100
11/11/54	14:15	100	100
11/11/54	14:20	100	100
11/11/54	14:25	100	100
11/11/54	14:30	100	100
11/11/54	14:35	100	100
11/11/54	14:40	100	100
11/11/54	14:45	100	100
11/11/54	14:50	100	100
11/11/54	14:55	100	100
11/11/54	15:00	100	100
11/11/54	15:05	100	100
11/11/54	15:10	100	100
11/11/54	15:15	100	100
11/11/54	15:20	100	100
11/11/54	15:25	100	100
11/11/54	15:30	100	100
11/11/54	15:35	100	100
11/11/54	15:40	100	100
11/11/54	15:45	100	100
11/11/54	15:50	100	100
11/11/54	15:55	100	100
11/11/54	16:00	100	100
11/11/54	16:05	100	100
11/11/54	16:10	100	100
11/11/54	16:15	100	100
11/11/54	16:20	100	100
11/11/54	16:25	100	100
11/11/54	16:30	100	100
11/11/54	16:35	100	100
11/11/54	16:40	100	100
11/11/54	16:45	100	100
11/11/54	16:50	100	100
11/11/54	16:55	100	100
11/11/54	17:00	100	100
11/11/54	17:05	100	100
11/11/54	17:10	100	100
11/11/54	17:15	100	100
11/11/54	17:20	100	100
11/11/54	17:25	100	100
11/11/54	17:30	100	100
11/11/54	17:35	100	100
11/11/54	17:40	100	100</

Query Ma
Best Loc
Matches

Q ₁	1
D ₅	499
Q ₂	51
D ₄	439
Q ₃	121
D ₆	379

RESULT 14
AGOS4901
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE AUTHORS	TITLE JOURNAL	REFERENCE AUTHORS	TITLE JOURNAL
----------------------	------------------	----------------------	------------------

EST.
Japanese medaka.
Oryzias latipes.
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Actinopterygii: Neopterygii: Teleostei: Euteleostei: Neoteleostei:
Acanthomorpha: Acanthopterygii: Percormorpha: Atherinomorpha:
Belontiiformes: Adrianichthyidae: Oryziinae: Oryzias.
1 (bases 1 to 535)
Naruse, K., Mitani, H. and Tanaka, M.
Medaka EST Project in University of Tokyo (2001)
unpublished (2001)
Contact: Kiyoshi Naruse
Department of Biological Sciences
Graduate School of Science, University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-3-5841-4443
Fax: 81-3-5841-4410
Email: naruse@btol.s.u-tokyo.ac.jp
This clone was isolated from OLHNI cell line cDNA library (OLb) 5
end sequences.

```

FEATURES:
end sequence
location/Qualifiers
    1..535
        /organism="Oryzias latipes"
        /strain="HNI"
        /db_xref="taxon:8090"
        /clone="OLB01.01h"
        /clone_lib="OLBNI cell line cDNA library (OLB)"
    150..135
        /size-count

```

Query Match	6.8%	Score 92.4:	DB 9:	Length 535:
Best Local Similarity	83.3%	Pred. No. 4.4e-06:		
Matches 105:	Conservative 0:	Mismatches 21:	Indels 0:	Gaps
Cy:	1	ccaatgaacgcacgtaatacgaactcactataggcgcaattgggtaccggggccccccctcg	60	
Dh	499	CCAGT GAGCGCGGGTAATACGACT CTACTATAGGGCGAAT TGGTACCGGGCCCCCCTCG	440	
Cy:	51	agcgccgcgcgccttt	120	
Dh	439	AGCGCGCGGGACCGTTTTTTTTTTTTTTTCTCATTTCAAGTGTGTTTTATTGGCA	380	
Cy:	121	tgaaaa	126	
Dh	379	GIACAA	374	

RESULT	14
AC054901	
LENGTH	658 bp
DEFINITION	Pan troglodytes DNA, clone: PTB-040L14.F, genomic survey sequence
ACCESSION	AC054901
VERSION	AC054901.1 GI:16592344
KEYWORDS	GSS: GSS (genome survey sequence).
SOURCE	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Ma BAC Library clone:PTB-040L14.F.
ORGANISM	Pan troglodytes
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan.

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (sites)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
BAC end sequences of Library PTB
unpublished
2 (bases 1 to 658)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chlimpbes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/)

COMMENT
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library. PB1 This BAC and
was generated during the R6d process and may have higher chance of
clone tracking errors.
DS1MPF5

PRIMERS
Sequencing: -21M13

```

Vector      : PKS145
R.Site 1   : Sac1
R.Site 2   : Sac1.
Location/Qualifiers
1. .658
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PIB-040L14.F"
/sex="male"
/cell_type="lymphoblast"
+PIB_Chimpanzee

```

BASE COUNT	184 a	157 c	127 g	189 t	1 others
ORIGIN					

Query Match	5.7%	Score 92;	DB 12;	Length 658;
Best Local Similarity	83.7%	Pred. No. 4.7e-06;		
Matches 144;	Conservative	0;	Mismatches 15;	Indels 13; Gaps 3;

Qy- 82 ttttttattgatcaagaattcaggcttttatattgagcaatgaaaacagctaaacttaatt 141
 |||||
 497 tttttctatttcacagantcacggctttatttgagcaatgaaaacagctaaacttaatt 546

DB 46/ 1111C1A11G181CAGAA111CAGGCT1A111A11A11G181CAGAA11CAG 193

Qy 142 tccaagcatgtgtagtcaaaagtctgc
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

Db 547 TCCAAGCATGTGTAGTTAAAGTATGCATGCGCACGAGGGGAGATATGGTTCACAAAACCA

Qy 194 at-tcaatggtttaaacac- - -tatttatttgagaacaaatatatttaa 240

Db 607 ATNTCAATGGTTAAACACATGGTTTATTTATGAGAACAAATATATTAAA 658

RESULT 15

AW790437

LOCUS AW790437 677 bp mRNA
DEFINITION C01755-F Lambda Zap, Stratagene Blumeria graminis f. sp. hordei
cdna clone C01755 similar to catalase, mRNA sequence
EST 01-MAY-2000

AW790437	GI:13902034
AW790437.1	
ACCESSION	
VERSION	

KEYWORDS
EST.
Blumeria graminis f. sp. hordei.
SOURCE
Blumeria graminis f. sp. hordei.
ORGANISM
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes; Erysiphales; Erysiphaceae; Blumeria.

REFERENCE
I (bases 1 to 677)
Thomas, S.W., Rasmussen, S.W., Glaring, M.A., Rouser, J.A. and Oliver, R.P.

TITLE Gene identification in the fungal pathogen *Blumeria graminis* by expressed sequence tag analysis

JOURNAL
Unpublished (2000)
 Contact: Paemissen S W

CONTACT: KASHIMIROV, S.M.
Department of Yeast Genetics

Carlsberg Laboratory
10 Gl. Carlsbergvej, DK-2500, Copenhagen, Denmark

Tel: 45 3327 5230
Fax: 45 3327 4755

Email: swr@crc.dk
High quality sequence stop: 677

FEATURES	Location/Qualifiers
POLYA=No.	

source 1.677

```
/organism="Blumella giannini"
/db_xref="taxon:62588"
```

```
/clone="CUI/55"  
/clone_lib="Lambda zap, Stratagene"
```

```

,
/cell_type="conidia"

```

```

BASE COUNT      /lab_host="Hordeum vulgare"
ORIGIN          162 a      179 c      191 g      143 t
                2 others

```

Query Match 6.5%; Score 88.8; DB 9; Length 677;
Best Local Similarity 93.0%; Pred. No. 1.6e-05;
Matches 93; Conservative 0; Mismatches 7; Indels

[illegible]

DB 19 GCCCGGGGATCCAC TAGTTCAGAGCGCCCCCCCCCCCCCCCCCCCCCCC

Qy: 1327 tccctttagtgaagggttaatttgcgcgccttggcgtatctcal 1368

Db 79 TCCCTTTAGTGAGGGTTAATTTCCAGCTTGGCGTAATCAT 118

Search completed: April 29, 2002, 17:16:55
Job time: 13214 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2002, 17:16:55 ; Search time 6783.7 Seconds
(without alignments)
1838.405 Million cell updates/sec

Title: US-09-248-178-52
Perfect score: 924
Sequence: 1 caaaggnacaggacagcgtt.....aaggcaaaaaaaaaaaaaa 924

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	505.6	54.7	519	9	A1917901
C 2	453	49.0	468	9	AW469374
C 3	449.2	48.6	461	9	A1801231
C 4	340.4	36.8	450	10	R78938
C 5	323	35.0	505	10	BG007631
C 6	258	27.9	414	9	AW070731
C 7	98.2	10.6	215	9	AA228748
C 8	61.2	6.6	321	10	BG009951
C 9	59.6	6.5	679	12	AG057824
C 10	58.4	6.3	393	9	A1937404
C 11	57.8	6.3	433	12	AQ115887
C 12	57.6	6.2	508	12	AQ598884
C 13	57.2	6.2	420	12	AQ153688
C 14	56.8	6.1	304	9	AA621843
C 15	56.8	6.1	372	9	A1056137
C 16	56.8	6.1	518	12	AQ885032
C 17	56.6	6.1	462	12	AQ226207

18	56.4	6.1	459	12	AQ210411
C 19	56.2	6.1	481	12	AQ001555
C 20	55.8	6.0	597	12	AQ554308
C 21	55	6.0	658	12	AG106298
C 22	54.8	5.9	360	9	A1081708
C 23	54.8	5.9	400	12	AQ789932
C 24	54.8	5.9	410	10	N88393
C 25	54.6	5.9	355	12	AQ051860
C 26	54.6	5.9	467	12	AQ580989
C 27	54.6	5.9	524	12	AQ695826
C 28	54.4	5.9	453	12	AQ129634
C 29	54.2	5.9	624	9	AW004000
C 30	54.2	5.9	842	12	AQ781743
C 31	53.8	5.8	754	12	AQ636028
C 32	53.8	5.8	869	10	BM007637
C 33	53.6	5.8	390	9	A1269041
C 34	53.6	5.8	392	9	AA463632
C 35	53.6	5.8	472	9	A1458879
C 36	53.6	5.8	480	9	A1580056
C 37	53.6	5.8	522	10	BE677940
C 38	53.6	5.8	561	10	BF196874
C 39	53.6	5.8	698	12	AG093829
C 40	53.2	5.8	440	12	AQ167953
C 41	53.2	5.8	999	9	AL547605
C 42	52.8	5.7	471	12	AQ123040
C 43	52.8	5.7	651	12	AG035074
C 44	52.4	5.7	503	12	AQ568884
C 45	52.4	5.7	538	9	AL042165

ALIGNMENTS

RESULT 1
A1917901/c
LOCUS tz14d04.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:228551 3',
DEFINITION mRNA sequence.
ACCESSION A1917901
VERSION A1917901.1 GI:5637756
KEYWORDS EST.
SOURCE human;
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 519)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CNA Library Preparation: Life Technologies, Inc.
CNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -400P from Gibco
High quality sequence stop: 397.
Location/Qualifiers
1. 519
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:228551"
/clone_lib="NCI_CGAP_Ut2"
/tissue_type="moderately-differentiated endometrial adenocarcinoma, 3 pooled tumors"
/lab_host="PH10B"
/note="organ: uterus; vector: pCMV-SPORT6; Site:1; Salt: Site_2: Not1; Cloned unidirectionally. Primer: Oligo dr.

Average insert size 1.85 kb. Life Technologies catalog #:

11539-012" 154 a 104 c 107 g 154 t

BASE COUNT
ORIGIN

Query Match 54.7%; Score 505.6; DB 9; Length 519;
Best Local Similarity 99.2%; Pred. No. 1.1e-103;
Matches 516; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Oy 391 aaatagctacttactattgtctattatttactttgtttgaaagctcaacaata 450
|||||
Db 519 AAATATAGTACTTACTATGTGCTATTTATTTACTTGTTCGAAGCTTAAGACCTCACAATA 460
Oy 451 gaatcccatcagccaccagacagagytgtgagtttctagtttggaaagctattaaa 510
|||||
Db 459 GAATCCCAATCAGCCACCAGACAGAGCTCTGAGTTTTCAGTTTGGAAAGAGCTATTAAA 490
Oy 511 taacaactctagtgctcaattctacttctgttctggttcaagtaactgggtcagcatttt 570
|||||
Db 399 TAACAACGCTAGTGTCAATTTCTATCTGTGTATGGTCAAGT-ACTGGGCTCAGCATTTT 341
Oy 571 acattcatgtctctttaaagttctagcaatgtgaagcaggaactatgatatttgacta 530
|||||
Db 340 ACATTCATTGTCTCTTTAAAGTTCTAGCAATGTGAAGCAGGAACATATGATTATATGACTA 281
Oy 631 cataaatgaagaaattgaggtctagatatacattaaatcttcccagggtcacacagcta 590
|||||
Db 280 CATAAATCAAGAAATGAGGCTCAGATACATTAAATTTCTCCAGGGTCACACAGCTA 221
Oy 691 gaactggcaagcctggagattgacctatccatgattccagcattggaagaatcataaatata 750
|||||
Db 220 GAACCTGGCAAGCCTGGGATTGATCCAGATCTTCCAGCATTTGAAGAAATCAATAATGTAA 161
Oy 751 ataactcgaagccttttccctcagaagagctcctggtgcttgctgacccaaccactagcact 810
|||||
Db 160 ATAACCTGCAAGCCCTTTTCCCTCAGAGAGCTCCTGGTGTTCACCAACCCACTAGCACT 101
Oy 811 tgttctcacagggggaacatctgtggcctgggaatcaactgcacgtcgcgaagagatttg 870
|||||
Db 100 TGTCTCTACAGGGGAAGATCTGTGGGCTGGGAATCAGTCAGCTGCGAAGAGATGTG 41
Oy 871 ctcttgatgaattattgttctcgtcagtggtggaaggca 910
|||||
Db 40 CTTCTGATGAATTATGTTCTCTGTCAGTGGTGTGAAGGCA 1

RESULT 2
AW469374/C
LOCUS hc82h01.x1 NCI_CGAP_Gas4 468 bp mRNA linear EST 24-FEB-2000
DEFINITION mRNA sequence.
ACCESSION AW469374
VERSION AW469374.1 GI:7039480
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 468)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapb-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: 40UP from Gibco
High quality sequence stop: 420.
Location/Qualifiers

FEATURES
source

1..468

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2899249"
/clone_lib="NCI_CGAP_Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signed ring cell features"
/lab_host="DH10B"
/note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.69 kb. Life Technologies catalog #:
11549-011"

BASE COUNT 133 a 98 c 103 g 134 t
ORIGIN

Query Match 49.0%; Score 453; DB 9; Length 468;
Best Local Similarity 98.9%; Pred. No. 7.7e-92;
Matches 464; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Oy 441 cctcaacatagaatcccatccagccaccagacagagagctgaggtttcttagtttgggaag 500
|||||
Db 468 CCTCACAATAGAAATCCCATCCAGCCACCAGACAGAGCTCTGAGTTTCTAGTTTGGGAAG 409
Oy 501 agctataaataaacaackttcagttcgaattctacttcttctatgctcaagtaactgggc 560
|||||
Db 408 AGCTATTAAATACACAGCTAGTGTCAATTTCTATCTATCTATGTTATGTTCAAGT-CTGGGC 350
Oy 561 tcagcattttacattcattgtctctttagtttctagcaatgtgaagcaggaactatgatt 620
|||||
Db 349 TCAGCAATTTTACATTCATTGCCTCTTTAAAGTTCTAGCAATGTGAAGCAGGAACACTATGAT 290
Oy 521 atattgactacataaataagaataattgaggtctcagatatacattaaatctctccagggt 680
|||||
Db 289 ATATTGACTACATAAATGAAGAAATTTGAGGCTCAGATACATTAAGTAAATCTCCAGGGT 230
Oy 581 cacacagctagaaactggcaagcctgggattgacatgatcttccagcattgaagaatc 740
|||||
Db 229 CACACAGCTAGAACTGGCAAGCCTGGGATTGATTCATGATCTTCCAGCATTTGAAGCATC 170
Oy 741 aataatgataaataactgcaagccttttctcagaagagctcctggtgcttgcaccaacc 800
|||||
Db 169 ATAAATGTAAATAACTGCAAGGCTTTTCTCTCAGAAAGAGCTCCTGTGTCTGCACCAACC 110
Oy 801 cactagcactgttctctacagggggaacatctgtggcctgggaatcaactgcacgtcga 860
|||||
Db 109 CACTAGCACCTGTCTCTACAGGGGAACATCTGTGGCCTGGGAATCACTGCACGTCGCA 50
Oy 861 aagatgttctctcagtaattatgttctcgtcagtggtggaaggc 909
|||||
Db 49 AGAGATGTTGCTTCTGTATGAATTTATGTTCTCTGTCAGTGTGTGAAGGC 1

RESULT 3

AI801231/C

LOCUS to86d10.x1 NCI_CGAP_Gas4 461 bp mRNA linear EST 14-DEC-1999

DEFINITION mRNA sequence.

ACCESSION AI801231

VERSION AI801231.1 GI:5366703

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 461)

NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

TITLE Tumor Gene Index

RESULT	4
R78938/c	
LOCUS	
DEFINITION	R78938 450 bp mRNA linear EST 09-JUN-1995 y187e07.sl Soares placenta N20HP Homo sapiens cDNA clone IMAGE:146244.3'. mRNA sequence.

```

Db 208 CACACAGCTAGAACTGCAGAGCCCTGGGATTGATCCCATGATCTTCCACCATTTGAAGAA 149
Oy 740 cataaataaataaactgaagagcccttttccctcagaagagct-cttggtgcttgacacaa 798
|||||
Db 148 CATAAATGTAATAACTGCAAGGCGCTTTTCCCTCAGAAGAGCTCCCTGGTCTTGACCAA 89
|||||
Oy 799 ccactagacactgttctctacaggggaacatctgtggccctgggaatacactgcacgtcg 858
|||||
Db 88 CCACATAGCAGCTTGTCTCTACAGGGGAACATCTGTGGGCCCTGGGAATCACTGCAAGTGG 29
|||||
Oy 859 caagagatgttgcttctgtatgaattatt 885
|||||
Db 28 CAAGAGATGTGTCTCTGATGAATTAT 1
|||||

RESULT 5
BGO 7631 505 bp mRNA linear EST 24-JAN-2001
LOCUS QV4-CN0250-281100-608-f05 GN0250 Homo sapiens cDNA, mRNA sequence.
DEFINITION BG007631
ACCESSION BG007631
VERSION BG007631.1 GI:12452005
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 505)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovsky-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Balaz,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
PROC Natl. Acad. Sci. U.S.A. 97 (7), 3491-3495 (2000)
MOLFEEDLINE 20202663
CONTACT: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl-QV4t2-QV4-GN0250-
281100-608-f05&t3-2000-11-28&t4-1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 505.
Location/Qualifiers
1..505
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0250"
/der_stage="Adult"
/note="Organ: Placenta normal; Vector: puc18; Site_1: SmaI
; Site_2: SmaI; A mini-library was made by cloning
products derived from ORSIES PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the pUC 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT 166 a 101 c 96 g 142 t
ORIGIN

```

Query Match 35.0% Score 323; DB 10; Length 505;
 Est Local Similarity 85.1%; Pred. No. 1.4e-52;
 Matches 395; Conservative 7; Mismatches 55; Indels 7; Gaps 5;

```

Oy 1 caaaggnacagaacagactttonaagaagtactnccatnccctcagaggacacagcccttt 60
|||||
Db 49 CCAAAGGACAGGACAGAGCTTG-AAAGTACTTTCAT-CTTCTCGAGGGACACAGCCCTTT 105
|||||
Oy 61 gcctccaaagcaatagaaaatttlaaagatttncactgagaagggggnccacgtttnart 120
|||||
Db 107 GCCTCCAAAGCAATAGCAAAATTTAAAGATTT-CACTGAGAAGGGGCC--ACGTTTAAAG 163
|||||
Oy 121 tntnaatgntcargnanartncccttncaaatgncrntcnaactnactnactnactnactnact 180
|||||
Db 154 TTATAATGTC TAAGGAGAGAGTCACTTCAAAATGCGATCACAATACATAAGATTTGGG 223
|||||
Oy 181 tncggnrtncmgnactatntcaggtttgaaaaactggatctgcccacttatcagtttatgtg 240
|||||
Db 224 TTCCAGAGTTCAGA--CATATCTAGTGTGTTGAAAACCTGGATCTGCCACTTACCAGTTATGTG 281
|||||
Oy 241 accttaagaacacgcgttaatttctcagagcctcagtttctctctataagttggaggt 300
|||||
Db 282 ACCITTAAGAACTCCGTTAAATTTCTCAGAGCCTCAGTTTCTTGTCTATAGTTGGGAGT 341
|||||
Oy 301 aaatataactatcatcttttccaaaggttgatggaacattaatgagtgaaatgacag 360
|||||
Db 342 AATATTACTATCATTTTTTCCAGAGGATTGATGTGAACATTAATGAGTGAATGACAG 401
|||||
Oy 361 atgtgtatcatggttctcctaataaacaatcacaataatagtagtacttacttattgttattat 420
|||||
Db 402 ATGTGTAICATGTTCTCTAATAACATCCAAATATAGTACTTACTATTGTCATTTAT 461
|||||
Oy 421 tactgtttgaagctaaagacacctcacaatagaaatcccatccagc 464
|||||
Db 462 TACTTGTITGAGCTAAGACCTCACAATAGAAATCCCATCCAGC 505
|||||

RESULT 6
AW070731/c 414 bp mRNA linear EST 20-OCT-2000
LOCUS xa04f05.x1 Soares_NFL1.GBC.S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2567361 3', mRNA sequence.
ACCESSION AW070731
VERSION AW070731.1 GI:6025729
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 414)
AUTHORS MCI-CGAP http://www.ncbi.nlm.nih.gov/mcicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 1011 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 396.
Location/Qualifiers
1..414
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2567361"
/clone_lib="Soares_NFL1.GBC.S1"
/note="Organ: pooled; Vector: pT73b-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
MCI-CGAP.GCBL) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of

```


Sequencing: M13Rev
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.

FEATURES	
SOURCE	

```

1. 1679
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-04411.R"
/sex="male"
/cell_type="lymphoblast"
/clone_id="PTB Chimpanzee Male BAC Library"
204 a 116 c 159 g 199 t 1 others
BASE COUNT
ORIGIN

```

```

Query Match      6.5%:  Score 59.6:  DB 12:  Length 679:
Best Local Similarity 52.8%:
Matches 153; Conservative 0: Mismatches 134; Indels 3

```

559

DB 518 AIIAIAIAIATCICIANAGAGAGIAI IRIITTCCTCAACGCGCGTTCGGCGCTTC

QY 547 aaqctcagatacattaagtaattctccagggtcacacagctagaactggcaaagcctg 706

100

Db 558 GAGGCTCAGAGAAGTGGTGTGTGATCCTAGGTCCTAAGATGAGGAGCTG 437

707 gaattgaatccatgacttccaccatgaaqaatcataaatgtaaataactgcaaggcctt 766

Db 498 AAATTTAACTTTTCATTGATTTGCACAAAGTCAGCAACACCCCTACTCTGTAAAGACACI 439

767 +ccctcacaagaagctcctggtacttataccacacccactagccacttqtctct 816

[illegible]

Db 438 TTCITAAAGACAGCAACTCA GTTATATTCATCCCTCACACCTCCTTCT 389

RESULT 10
AI937404/C

AI937404/c	AI937404	393 bp	linear	EST 08-MAR-2000
LOCUS	WP76905.x1	NCI_CGAP_Brn25	Homo sapiens	CDNA clone
DEFINITION	mRNA sequence. IMAGE:2467736 3',			

ACCESSION	AI937404
VERSION	AI937404.1
	GI:5676274

KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Mammalia; Primates; Hominidae; Homininae; Homo

REFERENCE

AUTHORS: NCI/NINDS-CGAP <http://>
TITLE National Cancer Instit
Disorders and Stroke

TRANAY
Disorders and Stroke,
(CCAP/BIGAP), Tumor Ge
Unpublished (1998)

JOURNAL
COMMENT
Unpublished (1998)
Contact: Robert Straus
Email: csapbs-r@mail.nyu.edu

English: Cúpulos (Cupules)
Tissue Procurement: Da
ph D

Ph.D.
cDNA Library Preparation
Bonaldo, Ph.D.

CDNA Library Arrayed
DNA Sequencing by: Wang

Clone distribution: N
found through the I.M.

www-bio.llnl.gov/bbrp/
Insert Length: 945

Seq primer: -40UP from

[illegible]

;

```

Query Match      6.1%; Score 56.8; DB 9: Length 304;
Best Local Similarity 61.7%; Pred. No. 0.012;
Matches 82; Conservative 2; Mismatches 49; Indels 0; Gaps 0;

Oy 180 tttccgrrtncmgnactatntcaggtttgaaaaactgagatctccacttctcagttatgt 239
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13 tttatagattgagcagatgctgggttcaaaagctctggttcttttcttcttggatgtgt 72
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 240 gacctaaagactccgcttaattctcagagctcagtttcccttgcctataagttggag 299
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 73 gacittgagcaagtacttacttctctgggctcacttttcccttctataaataaggaa 132
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 300 taatattaact 312
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133 taataatagtact 145
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
A1056137/c
LOCUS A1056137 372 bp mRNA linear EST 28-AUG-1998
DEFINITION ow33e08.s1 Soares_parathyroid_tumor_MbHPA Homo sapiens cDNA clone
IMAGE:1648646 3', mRNA sequence.
ACCESSION A1056137
VERSION A1056137.1 GI:3330003
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 372)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 455 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 365.
Location/Qualifiers
1..372
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1648646"
/clone_lib="Soares_parathyroid_tumor_MbHPA"
/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/notes="organ: parathyroid gland; vector: pT73D (Pharmacia
) with a modified polylinker; Site_1: Not 1; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer
15'-TGTTACCAACTGAAGTGGGCGCGCCACCAATTTTTTTTTTTTTTTT
TTTT-3'}. double-stranded cDNA was size selected, ligated
to Eco RI adapters (Pharmacia) digested with Not I and
cloned into the Not I and Eco RI sites of a modified pT73
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
adenomas was kindly provided by Dr. Stephen Marx, National
Institute of Diabetes and Digestive and Kidney Diseases.
NIH."
BASE COUNT 128 a 50 c 58 g 115 t
ORIGIN

```

```

Query Match      6.1%; Score 56.8; DB 9: Length 372;
Best Local Similarity 57.7%; Pred. No. 0.012;
Matches 120; Conservative 0; Mismatches 87; Indels 1; Gaps 1;

Oy 224 cacttatcagttatgtgaccttaaaagaactccgittaatttctcagagcctcagtttccct 283
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 334 CCCTCAGTAGTCACGTGTTCTTGGGTAAAGTCTTAACTTTTCTGTATCTCAGTTTCCTC 275
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 284 gtcataaacttgggagtaataattataactatcatcttttccaaaggattgatgtaaacatta 343
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 274 ATCTCTAAATAAGTAGGAATATTAAATAGTCT-ACACCTCATAAAGATTGTTGTGAAGATTA 216
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 344 atgaagtgaaatgacagatgtgtatcatgttctcctaataaacatccaaaatatagtactt 403
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215 AACGAGTTAATAAAGTGTGTGAATGATGCTCGGCACAAAGAAATCCACTCAATAAATAGT 156
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 404 actattgtcattattattacttctgttga 431
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155 AGTTCITATAGTTGTTGTATATGCTTTA 128
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: April 29, 2002, 17:17:00
Job time: 13219 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic acid nucleic search, using sw model

Run on: April 29, 2002, 19:41:12 : Search time 8020.54 Seconds
(without alignments)
2410.795 Million cell updates/sec

Title: US-09-248-178-62
Perfect score: 924
Sequence: 1 caaagacagacagctt.....aaggcaaaaaaaaaaaaaa 924

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10453268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1:	gb_ba.*	GenEmbl.*
2:	gb_hgt.*	
3:	gb_in.*	
4:	gb_om.*	
5:	gb_ov.*	
6:	gb_pat.*	
7:	gb_ph.*	
8:	gb_pl.*	
9:	gb_pi.*	
10:	gb_ro.*	
11:	gb_sts.*	
12:	gb_sy.*	
13:	gb_un.*	
14:	gb_vl.*	
15:	em_ba.*	
16:	em_fun.*	
17:	em_hum.*	
18:	em_in.*	
19:	em_mu.*	
20:	em_om.*	
21:	em_or.*	
22:	em_ov.*	
23:	em_pat.*	
24:	em_ph.*	
25:	em_pl.*	
26:	em_ro.*	
27:	em_sts.*	
28:	em_un.*	
29:	em_vl.*	
30:	em_hgt_hum.*	
31:	em_hgt_inv.*	
32:	em_hgt_other.*	
33:	em_hgtgo_inv.*	

ALIGNMENTS

RESULT	1
AC015461	
LOCUS	AC015461 Homo sapiens clone RP11-115N17, WORKING DRAFT SEQUENCE, 10
DEFINITION	AC015461 Homo sapiens clone RP11-115N17, WORKING DRAFT SEQUENCE, 10

ACCESSION	AC015461
VERSION	AC015461.3 GI:9502445
KEYWORDS	HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 158195)
JOURNAL	Birren, B., Linton, L., Nusbaum, C. and Lander, E.
REFERENCE	2 (bases 1 to 158195)
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

REFERENCE	Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhalter, B.,
AUTHORS	Brook, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
TITLE	Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donegan, L., Doyle, M.,
JOURNAL	Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
REFERENCE	Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Match	Length	DB	ID	Description
No.						

Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehoczy, J., Lie, C., Locke, K., Macdonald, P., Marquis, N., McEvan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C.H., O'Connor, J., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (16-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 27, 2000 this sequence version replaced gi:9112670.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WILBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1907
Center clone name: L15.N.17
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.9607.1
Consensus quality: 152466 bases at least Q40
Consensus quality: 154897 bases at least Q30
Consensus quality: 156006 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 157295; sum-of-contigs
Quality coverage: 5.8 in Q20 bases; agarose-fp
Quality coverage: 5.8 in Q20 bases; sum-of-contigs

***** NOTE: This is a 'working draft' sequence. It currently

consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1002: contig of 1002 bp in length
1003 1102: gap of 100 bp
1103 2388: contig of 1285 bp in length
2389 2489: gap of 100 bp
2489 3525: contig of 1037 bp in length
3525 3625: gap of 100 bp
3625 16527: contig of 12902 bp in length
16528 16527: gap of 100 bp
16528 28536: contig of 12009 bp in length
28537 28736: gap of 100 bp
28737 45598: contig of 17852 bp in length
45599 45698: gap of 100 bp
46599 71912: contig of 25214 bp in length
71913 72012: gap of 100 bp
72013 96612: contig of 24500 bp in length
96613 96712: gap of 100 bp
96713 128331: contig of 31619 bp in length
128332 128431: gap of 100 bp
128432 158195: contig of 29764 bp in length.

FEATURES

Source

1. 158195
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-115H17"
/clone_lib="RPC1-11 Human Male BAC"

misc_feature

1. 1002

/note="assembly_fragment"

misc_feature

1103. 2388

/note="assembly_fragment"

misc_feature

2489. 3525

/note="assembly_fragment"
3625. 16527
/note="assembly_fragment"
16628. 28636
/note="assembly_fragment"
clone_end:17
vector_side:right
28737. 46598
/note="assembly_fragment"
46699. 71912
/note="assembly_fragment"
72013. 96612
/note="assembly_fragment"
96713. 128331
/note="assembly_fragment"
128432. 158195
/note="assembly_fragment"
clone_end:SP6
vector_side:left
BASE COUNT 48628 a 30917 c 30188 g 47557 t 905 others
OPIGIN

Query Match 83.1% Score 768.2; DB 2; Length 158195;
Best Local Similarity 92.2%; Pred. No. 4.6e-178;
Matches 850; Conservative 9; Mismatches 55; Indels 8; Gaps 6;

OY 1 caaaggaagagaacagcttgaagtagtactgncatcncctnctcagagaccagcccttt 60
DB 148962 CCAAAAGGACAGGAACAGCTTG-AAAGTACTTGTCAT-CCITCTGCGAGGACCGCCCTTT 149019
OY 61 gcctccaaagcaataggaatttaaaagatttncactgagaaggggncacgtttnart 120
DB 149020 GCCTCCAAAGCAATAGGAATTTAAAGCATTT-CACGTGAGAAGGGGCC--ACGTTTAAAG 149076
OY 121 tntnaatgtntcargnarnartnccntncaaatgncrntcncactnactnrgnatgggt 180
DB 149077 TTATAAIGTCTAAGGAGAAGTCACTTCAAAATGCCAGCATCACATACAAATAGGAATTTGGG 149136
OY 181 tcccrrtncmactatnctcaggtttgaaaactggatcctccacttatccagtatgtg 240
DB 149137 TTCCAGAGTCAGA--CATACTAGGTTTGAAACTGGATCTGCCACTTATCATGTATGTG 149194
OY 241 accttaagaactccgttaatttctcagagcctcagtttccctgtctctataaattggaggt 300
DB 149195 ACCTTAAGAAGACTCCGTTAAATTTCTCAGAGCCTCAGTTTCTCTTGTCTATAAGTTGGAGT 149254
OY 301 aatattactactatcatcttttccaagagattgatgaacattaatgaagtgaaatgacag 360
DB 149255 AATATTAACTACTATCAATTTTCCAGGATTTGATGTGAACATTAAAGGAGTGAATGACAG 149314
OY 361 atgtgtatcatgattctcctaataaacatccaaataatagttacttactattgtcattattat 420
DB 149315 ATGTGTATCATGGTTCTTAATAAACATCCAAATAATAGTACTTACTATTGTCTATTATTAT 149374
OY 421 tacttttgaagactaaagacctccaaatagaatccatccagccaccagcagagagctc 480
DB 149375 TACTTGTITGAAGCTAAGACCTTCACAAATAGAAATCCCATCCAGCCACCAGACAGAGCTC 149434
OY 481 tgaatttcttagtttgaagagagctattataaataacaaacktctagtgctcaattcttactatgt 540
DB 149434 TGAGTTTCTTAGTTTGAAGAGAGCTAATTAAATACACAGCTAGTGTCAATTCCTATATCTTG 149494
OY 541 ttatgtcaagtaactgggctcagcattttacattcattgtctcttcttaagttcttaagaat 600
DB 149495 TTATGGTCAAGT-ACITGGGCTCAGCATTTTACATTCATTGTCCTTTTAAGTTCTAGCAAT 149553
OY 601 gtaaacagagactatgattatttgaactacataataatgaagaataatgaggctcagatata 660
DB 149554 GTGAAGCAGGAACATATGATTTATTTGACTACATNAATGAGANAATTCAGGCTCAGATACA 149613
OY 661 ttaagtaattctccaggggtcacacagctagaactgaaagcctgggattgtatccatga 720
|||||

```

Db 149614 TTAAGTAATCTCCAGGTCACACAGTAGAAGTGGCAAGCCCTGGGATTCATCCAGCA 149673
Oy 721 tcttcagcattgaagaatcataaagtataaataaactgaagacccctttcttcagagagc 780
|||||
Db 149674 TCTTCAGCAATGAGAAATCATAAATGTAATACATGCAAGGCCCTTTCTTCAGAGAGC 149733
Oy 781 tcttggtgttcacacacccactagcactgtttcttcacaggggaacectgtgggacct 840
|||||
Db 149734 TCTGTGCTTGACCAAGCCCACTAGCACTTGCTCTACAGGGGAACAATGTGTGGGCT 149793
Oy 841 ggaatcactgcacatcacaagagatgttcttcagatgaattattgttctgtcagtg 900
|||||
Db 149794 GGAATCACTGACGTCGCAAGAGATGTGCTTCATGATTAATGTTCTGTCAGTGG 149853
Oy 901 tgtgaaggcaaaaaa 922
|||||
Db 149854 TGTAAGGCAAAAAA 149875

RESULT 2
AL157394/c
LOCUS
DEFINITION
  AL157394 187313 bp DNA linear PFI 22-AUG-2001
  Human DNA sequence from clone RP11-399019 on chr:mosome 10,
  complete sequence.
ACCESSION
  AL157394
VERSION
  AL157394.15 GI:15384622
KEYWORDS
  HTG.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 187313)
  Blakey,S.
  Direct Submission
  Submitted (22-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
  CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
  requests: clonerequest@sanger.ac.uk
  On Aug 31, 2001 this sequence version replaced gi:14161146.
  During sequence assembly data is compared from overlapping clones.
  Where differences are found these are annotated as variations
  together with a note of the overlapping clone name. Note that the
  variation annotation may not be found in the sequence submission
  corresponding to the overlapping clone, as we submit sequences with
  only a small overlap as described above.
  This sequence was finished as follows unless otherwise noted: all
  regions were either double-stranded or sequenced with an alternate
  chemistry or covered by high quality data (i.e., phred quality >=
  30); an attempt was made to resolve all sequencing problems, such
  as compressions and repeats; all regions were covered by at least
  one plasmid subclone or more than one M13 subclone; and the
  assembly was confirmed by restriction digest. The following
  abbreviations are used to associate primary accession numbers given
  in the feature table with their source databases: Emi., EMBL; S.,
  SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information on the WormPEP
  database can be found at
  http://www.sanger.ac.uk/Projects/C_elegans/wormpep
  This sequence
  was generated from part of bacterial clone contigs of human
  chromosome 10, constructed by the Sanger Centre Chromosome 10
  Mapping Group. Further information can be found at
  http://www.sanger.ac.uk/HGP/chr10
  RP11-399019 is from the library RP11-11.2 constructed by the group
  of Pieter de Jong. For further details see
  http://www.chori.org/bacpac/home.htm
  VECTIR: pBACE3.5
  This sequence is the entire insert of clone PFI-399019. The true
  left end of clone RP11-496H23 is at 166408 in this sequence. The
  true right end of clone RP11-30415 is at 18704 in this sequence.
  Location/Qualifiers
    1..187313
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="10"
      /clone="RP11-399019"

```

FEATURES

source

```

/misc_feature
  100119..100156
    /note="Sequence confirmed by AC015461 sequenced by WIBR."
/misc_feature
  100157..100198
    /note="Sequence from AC015461 sequenced by WIBR."
/misc_feature
  105808..105972
    /note="Sequence from AC015461 sequenced by WIBR."
/misc_feature
  105973..105989
    /note="Sequence confirmed by AC015461 sequenced by WIBR."
BASE COUNT 55669 a 36398 c 36888 g 58358 t
ORIGIN
Query Match      83.1%; Score 768.2; DB 9; Length 187313;
Best Local Similarity 92.2%; Pred. No. 4.5e-178;
Matches 850; Conservative 9; Mismatches 55; Indels 8; Gaps 4;
Oy 1 caaggnacaggaacagcttgnaaagtactuncatnccctnctcagagggagccgctttt 60
  ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 95976 CCAAGGACAGGACAGCTTG-AAAGTACTTGCAAT-CCTTCTCGAGGACCGCCCTTT 95919
Oy 51 gcttcacacacacacacacacacacacacacacacacacacacacacacacacacacac 120
  ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 96919 GCTTCACAAAGCAATAGGAAATTTAAAAGATTT-CACTGAGAAGGGGCC--ACGTTTAAG 96862
Oy 121 tntnaatgtnicargnananarntnccctncaaaatgncrncnctnactnactnrgnattgg 180
  ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 96861 TTATAATGTCTAAGGAGAGATCCTTCAAAATGCGAGCATCACAATACATAGGAATTCGG 96802
Oy 181 tncqnrntnncnactatntcaggtttgaaacactggatctgcccacttatcagttatdg 240
  ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 96801 TTCAGAGTCAGA--CATATCTAGGTTTGAACAACTGGATCTGCCACTTATCAGTTATGTG 96744
Oy 241 accttaagaaacccgcttaattctcagagccctcagtttcttctgtctataagttggaggt 300
  ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 96743 ACCTTAAAGAACTCCGTTAAATTTCTCAGAGCCCTCAGTTTCTCTGTCTATAAGTTGGGAGT 96684
Oy 301 aatataatactatcatcttttccaaaggattgagtgaacatttaagtgaagtgaaatgacag 360
  ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 96683 AATAATATACTATCATTTTTCAGAGGATTGATGTGAACATTAATGAGGTGAATGACAG 96624
Oy 361 atgtatatactggttcttaataaataccaaataatagtaacttacttacttacttactat 420
  ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 96623 ATGTGATATCATGGTTCCTCTAAATAAATCCAAATATAGTACTTACTATGTCTATTATAT 96564
Oy 421 tactgtttgaagctaaagacacctcacaatagaatccctccagccacacagagagctc 480
  ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 96563 TACTTGTITGAAGCTAAAGACCTTCACATAGAAATCCATCCAGCCACCCAGACAGAGCTC 96504
Oy 481 tgaagtttctagtttgaagagagctattaaatacaacactctagtgctcaattctacttg 540
  ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 96503 TCAGTTTCTAGTTTGGAGAGAGCTATTAAATAACAACGCTCTAGTGTCAATTTCTACTTTG 96444
Oy 541 ttatagtcagtaactagggctcagcattttacattcttctcttcttaagttctagcaat 600
  ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 96443 TTATGCTCAAGT-ACCTGGGCTCAGCAATTTTACATTCATTGTCTCTTTTAAGTTCTAGCAAT 96385
Oy 601 gtaagcagagaaactatgattatattgactacataaataagaataatgaggtctcagataca 660
  ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 96384 GTGAAGCAGGAACTATGATTATATTGACTTACATATAATGAAGAAATGAGGCTCAGATACA 96325
Oy 661 taagtaattctccagaggttcacacagctagaaactggcaaacctgggattatccatga 720
  ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 96324 TTAAGTAATCTCCAGGTCACACAGTAGAAGTGGCAAGCCCTGGGAATGATCATCCACA 96265
Oy 721 tcttcagcattgaagaatcataaataaataaactgcaagggccttttcttcagagagac 780
  ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 96264 TCTTCAGCAATTCAGAAATCATAAATGTAATACATGCAAGGCCCTTTCTTCAGAGAGAC 96205
Oy 781 tcttggtgttcacacacccactagcactgttctctcacaggggaacactctgtggacct 840
  ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 96204 TCTTGTGCTTGCAACCAAGCCCACTAGCACTTGTCTCTACAGGGGAACATCTGTGGGCT 96145

```

```
Oy 841 ggaatcactgcacgcacgaagatgttgccttcgataaattatgttctcgtcagtg 900
|||||
Db 96144 GGAATCAGTGCACGTCGCAAGAGATGTTGCTTCGATGAATAT:GTCCGTCAGTGG 36085
|||||

Oy 901 tgtgaaggcaaaaaa 922
|||||
Db 96084 IGTAAGCAAAAAA 95063
|||||

RESULT 3
G43743/c
LOCUS
DEFINITION
WIAF-3334-SIS Human Thudson SANGER Homo sapiens STS genomic,
sequence tagged site.
ACCESSION
G43743
VERSION
G43743.1 GI:4192660
KEYWORDS
SIS.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 214)
Wang,D.G., Fan,J.B., Siao,C.J., Berno,A., Young,F., Sapolsky,R.,
Ghandour,G., Perkins,N., Winchester,E., Spencer,J., Kruglyak,L.,
Stein,L., Hsieh,L., Topaloglou,T., Hubbell,E., Robinson,E.,
Mittmann,M., Morris,M.S., Shen,N., Kilburn,D., Rioux,J.,
Nusbaum,C., Rozen,S., Hudson,T.J., Lipshutz,R., Chee,M. and
Lander,E.S.
Large-scale identification, mapping, and genotyping of
single-nucleotide polymorphisms in the human genome
Science 280 (5365), 1077-1082 (1998)
98248615
9582121
Synonyms: stSG15752a, stSG15752
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: CATCAGAGCAACATCTCTTGC
Primer B: ATTCTCCAGGTCACACAG
STS size: 214
PCR Profile:
Presoak: 94 degrees C for 4.00 minutes
Denaturation: 94 degrees C for 50.0 seconds
Annealing: 58 degrees C for 1.50 minutes
Polymerization: 72 degrees C for 1.00 minutes
PCR Cycles: 30
Thermal Cycler: custom built by IAS, Costar, Cambridge MA

Protocol:
Template: 10 ng
Primer: each 5 pM
dNTPs: 4 mM
Taq Polymerase: 0.5 U
Total Vol: 20 uL

Buffer:
Mg2+: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
Gelatin: .001 %
Location/Qualifiers
1. .214
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="112.70 kb from top of Chr10 linkage group"
/clone_lib="Human Thudson SANGER"
/notes="human STS created from EST in the Sanger database"
1. .214
1. .22
complement(195..214)

STS
primer_bind
primer_bind
```

```
BASE COUNT 52 a 48 c 56 g 57 t 1 others
ORIGIN
Query Match 19.9% Score 184.2; DB 11; Length 214;
Best Local Similarity 97.2%; Pred. No. 9,1e-35;
Matches 208; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

Oy 568 attctccagggtcacacagctagaaactggcaaaag-cctgggattgacataatcttc 726
|||||
Db 214 ATTCTCCAGGTCACACAGCTAGAAACTGCAAGCCCTGGGATTGATCATGATCTNCC 155
|||||

Oy 727 agcattgaagaatcataaattgaataaactgcaagcctttctcctcagaagact-cctg 785
|||||
Db 154 AGCATTTGAAGAAICATAAATGTAATACATGCAAGGCTTTTCTCTCAGAGAGCTCCCTG 95
|||||

Oy 786 gtgcttgaccacacccactagcactgttctctacaggggaaacatctgtggcctcggaaa 845
|||||
Db 94 GTGCTTGCAACCAACCCACTAGCACTTGTTCTCTACAGGGAAACATCTGTGGCCCTGGAA 35
|||||

Oy 845 tcaatgcacgtcgaagagatgttgccttcgaatg 879
|||||
Db 34 TCACITGCACGTCGCAAGAGATGTTGCTTCTGATG 1
|||||

RESULT 4
G43744/c
LOCUS
DEFINITION
WIAF-3335-SIS Human Thudson SANGER Homo sapiens STS genomic,
sequence tagged site.
ACCESSION
G43744
VERSION
G43744.1 GI:4192661
KEYWORDS
SIS.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 214)
Wang,D.G., Fan,J.B., Siao,C.J., Berno,A., Young,P., Sapolsky,R.,
Ghandour,G., Perkins,N., Winchester,E., Spencer,J., Kruglyak,L.,
Stein,L., Hsieh,L., Topaloglou,T., Hubbell,E., Robinson,E.,
Mittmann,M., Morris,M.S., Shen,N., Kilburn,D., Rioux,J.,
Nusbaum,C., Rozen,S., Hudson,T.J., Lipshutz,R., Chee,M. and
Lander,E.S.
Large-scale identification, mapping, and genotyping of
single-nucleotide polymorphisms in the human genome
Science 280 (5366), 1077-1082 (1998)
98248615
9582121
Synonyms: stSG15752b, stSG15752
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: CATCAGAGCAACATCTCTTGC
Primer B: ATTCTCCAGGTCACACAG
STS size: 214
PCR Profile:
Presoak: 94 degrees C for 4.00 minutes
Denaturation: 94 degrees C for 50.0 seconds
Annealing: 58 degrees C for 1.50 minutes
Polymerization: 72 degrees C for 1.00 minutes
PCR Cycles: 30
Thermal Cycler: custom built by IAS, Costar, Cambridge MA

Protocol:
Template: 10 ng
Primer: each 5 pM
dNTPs: 4 mM
Taq Polymerase: 0.5 U
Total Vol: 20 uL
```

Buffer: Mg2+: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
Gelatin: .001 %
Location/Qualifiers
1. .214
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="112.70 CR from top of Chr10 linkage group"
/clone_lib="Human Thudson SANGER"
/note="human STS created from EST in the Sanger database"
1. .214
STS
primer_bind
primer_bind
BASE COUNT 52 a 48 c 56 g 57 t 1 others
ORIGIN
Query Match 19.9% Score 184.2; DB 11: Length 214;
Best Local Similarity 97.2%; Pred. No. 9.1e-35;
Matches 208; Conservative 0; Mismatches 4; Indels 2; Gaps 2;
Qy 668 attctccagggtcacacagctagactggcaag-cctgggattgatccatgatcttcc 726
Db 214 ATCTCCAGGGTCACACAGCTAGAACTGCAAGCCCTGGGATTGATCCAGATCTCC 155
Qy 727 agcattgaagaatcataatgataactgaagcagcgtcttctcagaagaact-ctg 785
Db 154 AGCATTCGAAGATCAATAATGATAAATACTGCAAGGCCCTTTTCCACGAAGAGCTCCCTG 95
Qy 786 gtactgcaccacccactagcactgttctctacaggggaaacatctgtggcctggaaa 845
Db 94 GTGCTTGACCAACCCACTAGCACATGTTCTACAGGGGAACATCTGTGGCCCTGGGAA 35
Qy 846 tcactgcacgtcgcaagagatgttgcctctgatg 879
Db 34 TCACCTGCACGTCGAAGAGATGTTGCTCTGTATG 1
RESULT 5
AC101755/c
LOCUS
DEFINITION Mus musculus clone RP24-390H23, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC101755
VERSION AC101755.1 GI:17060530
KEYWORDS HTG; HTGS PHASE0.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 56933)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP24-390H23
Unpublished
2 (bases 1 to 56933)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Labrecque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,N., Liu,G.,
MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J.,
Meneus,L., Mihoza,T., Mlenga,V., Murphy,I., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Olliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pillara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,

Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Toham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L17347
Center clone name: 390_H_23

* NOTE: This record contains 84 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 654: contig of 654 bp in length
555 754: gap of 100 bp
755 1446: contig of 692 bp in length
1447 1546: gap of 100 bp
1547 2238: contig of 692 bp in length
2239 2338: gap of 100 bp
2339 3017: contig of 679 bp in length
3018 3117: gap of 100 bp
3118 3798: contig of 681 bp in length
3799 3898: gap of 100 bp
3899 4590: contig of 692 bp in length
4591 4690: gap of 100 bp
4691 5366: contig of 676 bp in length
5367 5466: gap of 100 bp
5467 6131: contig of 667 bp in length
6134 6233: gap of 100 bp
6234 6927: contig of 694 bp in length
6928 7027: gap of 100 bp
7028 7721: contig of 694 bp in length
7722 7821: gap of 100 bp
7822 8524: contig of 703 bp in length
8525 8624: gap of 100 bp
8625 9303: contig of 679 bp in length
9304 9403: gap of 100 bp
9404 10099: contig of 696 bp in length
10100 10199: gap of 100 bp
10200 10903: contig of 704 bp in length
10904 11003: gap of 100 bp
11004 11671: contig of 668 bp in length
11672 11771: gap of 100 bp
11772 12472: contig of 701 bp in length
12473 12572: gap of 100 bp
12573 13261: contig of 689 bp in length
13262 13361: gap of 100 bp
13362 14063: contig of 702 bp in length
14064 14163: gap of 100 bp
14164 14843: contig of 680 bp in length
14844 14943: gap of 100 bp
14944 15445: contig of 702 bp in length
15445 15745: gap of 100 bp
15746 16452: contig of 707 bp in length

TITLE
JOURNAL
COMMENT

[illegible]

LOCUS HS57G9 113872 bp DNA linear PFI 12-DEC-1999
DEFINITION Human DNA sequence from clone CTA-57G9 on chromosome 22q12.1. Contains part of the gene for a novel Kringle and CUB domain protein, ESI5, STSS, CA repeat polymorphisms, GSSs and a CpG Island, complete sequence.
ACCESSION Z95116
VERSION Z95116.1 GI:2832592
KEYWORDS HTG: ca repeat polymorphism; CpG island; CUB domain; Kringle domain.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 113872)
AUTHORS McLaren, S.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT requests: clonerequest@sanger.ac.uk
 On Feb 5, 1998 this sequence version replaced gi:2578126.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>
 CTA-57G9 is from the human BAC library described in U-J. Kim et al. (1995) Genomics 34, 213-218.
FEATURES SOURCE
 Location/Qualifiers
 1..113872
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="22"
 /map="q12.1"
 /clone="CTA-57G9"
 /clone_lib="CIT978SK-A1"
 complement(1..284)
 /note="match: GSS: Em:B14184
 match: STS: Em:B14184"
 35..392
 /note="match: GSS: Em:B14225"
 complement(305..682)
 /note="match: GSS: Em:B14096"
 593..618
 /note="13 copies 2 mer at 96 conserved"
 740..1035
 /note="AluY repeat: matches 1..296 of consensus"
 1326..1841
 /note="MLT1E repeat: matches 44..559 of consensus"
 1857..2202
 /note="MLT1A2 repeat: matches 24..374 of consensus"
 2346..2519
 /note="MER21B repeat: matches 160..318 of consensus"
 2520..2665
 /note="FLAM_C repeat: matches 1..133 of consensus"

2666..3075
 /note="MER21B repeat: matches 318..792 of consensus"
 3241..3458
 /note="MLT1D repeat: matches 1..242 of consensus"
 4048..4345
 /note="MLT1D repeat: matches 206..505 of consensus"
 4665..4996
 /note="match: GSS: Em:A0119749"
 join(4811..4935,8717..8870,20796..21128,22078..22209)
 /gene="BK57G9.1"
 join(4811..4935,8717..8870,20796..21128,22078..22209)
 /gene="BK57G9.1"
 /note="supported by GENSCAN
 match: ESTs: Em:AA506741 Em:AA670682
 match: Proteins: Tr:P91972 Sw:P98065 Tr:O15113 Tr:O351113
 Tr:Q24132 Tr:O70244 Tr:O57433 Tr:O88204 Sw:P97435
 Tr:O60494 Tr:O57658 Tr:O57074 Tr:O57434 Tr:O08628
 Tr:O61398 Sw:P98066 Tr:O23995 Sw:P97333 Tr:O08859
 Tr:O95917"
 /codon_start=3
 /evidence="not experimental
 /product="BK57G9.1 (novel Kringle and CUB domain protein)"
 /protein_id="CA662952.1"
 /db_xref="GI:6572252"
 /translation="PCNLGKYKHGPNPPLTGTSKTSNKLTIQTICFSQRKRFAG
 MESGVACFCGNPNPDYKYGCAASTCNCVFGDHTPCGCGDRIILFDLVGACGNY
 SAMSVVAVSPDFDIYATGVCTYIRVPGASHIHFSPLFDIROSADVLLDGVTH
 RVLARFHGRSPRLSFNVSLDFVILYFESDRINQAGFAVLQYQVKEELPOERPAVNO
 TVAEVIIDQANLSVSAKSSKVLVYIITS"
 4957..5312
 /note="LIMB6 repeat: matches 5808..6175 of consensus"
 5309..6111
 /note="LIME1 repeat: matches 5016..5828 of consensus"
 6112..6422
 /note="AluSg repeat: matches 1..308 of consensus"
 6423..6546
 /note="LIME1 repeat: matches 5828..5952 of consensus"
 6502..6965
 /note="LIMD2 repeat: matches 4987..5472 of consensus"
 6966..7272
 /note="AluSx repeat: matches 3..311 of consensus"
 7273..7944
 /note="LIMD2 repeat: matches 5472..6069 of consensus"
 7945..8246
 /note="AluSx repeat: matches 1..302 of consensus"
 8247..8328
 /note="LIMD2 repeat: matches 6069..6153 of consensus"
 8368..8426
 /note="MIR repeat: matches 201..262 of consensus"
 8959..9050
 /note="L2 repeat: matches 2641..2732 of consensus"
 9708..10119
 /note="MLT1C repeat: matches 8..464 of consensus"
 10556..10724
 /note="MIR repeat: matches 74..252 of consensus"
 10783..10914
 /note="LIMB6 repeat: matches 5448..5581 of consensus"
 10915..11122
 /note="AluSg/x repeat: matches 81..296 of consensus"
 11123..11661
 /note="LIMB6 repeat: matches 5576..6122 of consensus"
 11710..11811
 /note="MER47 repeat: matches 2224..2323 of consensus"
 12111..12286
 /note="MIR repeat: matches 47..262 of consensus"
 12663..12810
 /note="AluSg/x repeat: matches 141..292 of consensus"
 13522..13690
 /note="MIR repeat: matches 78..252 of consensus"
 13684..13741
 /note="L2 repeat: matches 2608..2666 of consensus"
 14535..14774
 /note="MIR repeat: matches 18..259 of consensus"

repeat_region	14781..15016 /note="L1ME1 repeat: matches 5918..16175 of consensus" 15017..15226 /note="105 copies 2 mer ta 80 conserved" 15233..15319 /note="L1ME1 repeat: matches 5867..5952 of consensus"
repeat_region	15320..15625 /note="AluSp repeat: matches 1..307 of consensus" 15626..16135 /note="L1ME1 repeat: matches 5353..5867 of consensus"
repeat_region	16114..16420 /note="L1M4c repeat: matches 1692..2021 of consensus" 16421..16731 /note="AluY repeat: matches 1..311 of consensus"
repeat_region	16732..16757 /note="L1M4c repeat: matches 1656..1692 of consensus"
repeat_region	16758..17052 /note="AluJo repeat: matches 3..295 of consensus"
repeat_region	17053..17467 /note="L1M4c repeat: matches 1329..1665 of consensus"
repeat_region	17468..17572 /note="U6 repeat: matches 1..105 of consensus"
repeat_region	17603..17613 /note="L1M4c repeat: matches 1194..1204 of consensus"
repeat_region	17614..17910 /note="AluJo repeat: matches 1..298 of consensus"
repeat_region	17911..18080 /note="L1M4c repeat: matches 1024..1194 of consensus"
repeat_region	19378..19673 /note="AluSx repeat: matches 3..298 of consensus"
repeat_region	19746..20048 /note="AluY repeat: matches 1..301 of consensus"
repeat_region	21731..21824 /note="L2 repeat: matches 2355..2453 of consensus"
repeat_region	22594..22773 /note="L2 repeat: matches 2546..2750 of consensus"
repeat_region	22774..23077 /note="AluJo repeat: matches 1..303 of consensus"
repeat_region	23078..23176 /note="AluJo/FRAM repeat: matches 134..232 of consensus"
repeat_region	23367..23455 /note="MIR repeat: matches 58..144 of consensus"
misc_feature	complement(23800..24202) /note="match: GSS: Em:A0183044"
misc_feature	complement(24577..24929) /note="match: SIS: Em:Z565939 match: SIS: Em:HS284VA9"
repeat_region	24681..24774 /note="47 copies 2 mer gt 90 conserved"
repeat_region	26322..26634 /note="AluSg repeat: matches 1..313 of consensus"
repeat_region	26915..27209 /note="AluY repeat: matches 1..295 of consensus"
misc_feature	complement(29985..30269)

* be preserved.
* 1 67: contig of 67 bp in length
* 68 167: gap of 100 bp
* 168 1872: contig of 1705 bp in length
* 1873 1972: gap of 100 bp
* 1973 7959: contig of 5987 bp in length
* 7960 8059: gap of 100 bp
* 8060 14082: contig of 6023 bp in length
* 14083 14182: gap of 100 bp
* 14183 24347: contig of 10165 bp in length
* 24348 24447: gap of 100 bp
* 24448 44018: contig of 19571 bp in length
* 44019 44118: gap of 100 bp
* 44119 52009: contig of 17891 bp in length
* 52010 52103: gap of 100 bp
* 52110 84737: contig of 22628 bp in length
* 84738 84837: gap of 100 bp
* 84838 115856: contig of 31019 bp in length
* 115857 115957: gap of 100 bp
* 115957 162051: contig of 46105 bp in length.

FEATURES
source

Location/Qualifiers
1. 162051
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone_lib="RPC1-11 Human Male BAC"
/clone_end:T7
1. 67
/note="assembly_fragment"
vector_side:right
misc_feature 158..1872
/note="assembly_fragment"
1973..7959
misc_feature /note="assembly_fragment"
8060..14082
misc_feature /note="assembly_fragment"
clone_end:SP6
vector_side:left
14183..24347
/note="assembly_fragment"
24448..44018
misc_feature /note="assembly_fragment"
44119..52009
misc_feature /note="assembly_fragment"
52110..84737
misc_feature /note="assembly_fragment"
84838..115856
misc_feature /note="assembly_fragment"
115957..162051
/note="assembly_fragment"
60903 a 30373 c 31811 g 48073 t 901 others

BASE COUNT 50903 a 30373 c 31811 g 48073 t
ORIGIN

Query Match 7.3%; Score 67.4; DB 2: Length 162051;
Best Local Similarity 56.4%; Pred. No. 1.8e-06;
Matches 145; Conservative 1; Mismatches 107; Indels 5; Gaps 1:
Oy 502 gctattaataacaacttagtgcattctctattgtatggtcaagtaactgcacct 561
Db 47111 GTTACTAACCCACTAATGCTAACATTTATTAACTATTATTATGCTAGGCACGTCTCT 47170
Oy 562 cagcattttacattctctctcttttaagttcttagcaaa-----tgtgaagcaggaaactat 616
Db 47171 TAGCACTTGAGATAAGCACTTTCATCCCAATTTTTCATATATCTATATAGGAAGCACTAT 47230
Oy 517 gattattatgaactacataaataaataatgagggctcagatcatatgaatattctccca 676
Db 47231 TATTATCAATCATTTTCAAAATAAAGATAATGAGGCACAGAGATTAGTAAATTCGCCCA 47250
Oy 577 ggggtcacacagctagaaactggcaagcctgggattgattccatcatcttccacattgaag 736

Db 47291 AGGTCACACAGCTGTGAAGTGACAGAGAGGAGGATTTTAGCCAGGCACTGTGGTTTCAGA 47350
Oy 737 aatcataaatgtataataac 755
Db 47351 GCATACATACCATTAATTACC 47369

RESULT 11
LOCUS

AL133551 175940 bp DNA linear PRI 21-AUG-2000
Human DNA sequence from clone RP11-57G10 on chromosome 10. Contains
a J-domain containing protein (JDP1) isoform B, the SIRT1 gene
(Sirt2-like proteins (sirtuins) type 1), part of a novel gene
similar to KIAA0032, two ribosomal pseudogenes, 2 CpG islands,
ESTs, STSS and GSSs, complete sequence.
AL133551
VERSION AL133551.13 GI:9407715
KEYWORDS HIT: J-domain; RPL12; RPL21; SIRT1.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 175940)
/map="3"
Direct Submission
Submitted (21-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 23, 2000 this sequence version replaced gi:9367364.

COMMENT

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/Celegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-57G10 is from the library RPC1-11.1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pBACE3.6
This sequence is the entire insert of clone RP11-57G10 The true
left end of clone RP11-474D14 is at 172652 in this sequence.

FEATURES

Source

Location/Qualifiers
1..175940
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-57G10"
/clone_lib="RPC1-11.1"
1..2318
/note="LIPAL0 repeat: matches 3815. .6158 of consensus"
2338..2767
repeat_region
/note="LIMA4 repeat: matches 4946. .5385 of consensus"
2768..3058
repeat_region
/note="AluJo repeat: matches 1. .296 of consensus"
3059..3371
repeat_region
/note="LIMA4 repeat: matches 5385. .5704 of consensus"
3372..3682
repeat_region

Qy	507	taaatacaacackcttagtgtaaatcttataactgtgtatggccaagtaactggcgctcagca	566	
Db	12394	TCACCACCACCATCTAATGTTGATTGAGCCCTTGCATGTGCCAGCTTCTGCTACTGAGCT	12453	
Qy	567	ttttacattcatttctctttaaagttcttagca	-----atgtgaagcaggaacatagatt	620
Db	12454	CITTGCAATGGAATATCTAATTTAATCTTCACACCACCTATTTGAAGTAAGTACTATTAAT	12513	
Qy	621	atattgactacataaaatgaagaatattgagcttcagatcacattaaagtaattctccaggg	680	
Db	12514	AATCTCATTTTATAGAGGAGGAACTGAGGTTTAGAGAGGTTTAAATAATTTGCCCAAGG	12573	
Qy	681	cacacagctagaactggcaagcctgggatttgcattgcattccagcattggaagatc	740	
Db	12574	CATAAGCTAGTAAGTACTGAACTGGGAATTTGAATTTTAAACCTGGCGAGCTGACTGCCAA	12633	
Qy	741	ataaatgttaataaactcgaagccttttccctcagaagaagcctctgggtgctt	791	
Db	12634	AGCCAGCTACTTAACTGCTGCACAGTACTTCCAAAGGCCTGTCTGCAGCTT	12684	
RESULT	13			
AC078778				
LOCUS		207856 bp	DNA linear	
DEFINITION		Homo sapiens chromosome 12 clone RP11-968A15, WORKING DRAFT	HTG 25-JUN-2001	
ACCESSION		SEQUENCE, 8 unordered pieces.		
VERSION		AC078778.23	GI:14547475	
KEYWORDS		HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE		human.		
ORGANISM		Homo sapiens		
Eukaryota:		Chordata:	Craniata:	
Mammalia:		Eutheria:	Primates:	
			Catarrhini:	
			Hominidae:	
			Homo:	
			1 (bases 1 to 207856)	
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,				
Aisbrooks,S.L., Amarakunte,H.C., Are,J.R., Banks,T., Barbara,J.				
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,				
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,				
Hurch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,				
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,				
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,D.B., Cox,C.,				
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,				
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,				
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,				
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,				
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,				
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,				
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,				
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,				
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,				
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,				
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,				
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,				
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,				
Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,				
Louised,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,				
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,				
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,				
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K.,				
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,				
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkwo,S.,				
Ogulu,M., Okunnu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,				
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,				
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojokhan,I., Rolfe,M.,				
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shookster,N.,				
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,				
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tameria,A., Tameria,K.,				
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,				
Thomas,S., Usmani,K., Vasquez,L., Vera,Y., Villafan,D., Vinson,R.,				
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,				
Washington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,				
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,J., Nelson,D.,				

Thomas, S. Y., Osmani, K. A., Buescher, B., Kera, S., and Porter, C. 1993. Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, K. R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorilla, S., Nelson, D.

Matches	102:	Conservative	0:	Mismatches	51:	Indels	0:	Gaps	0:
Qy	529	attctatctattatgtcgaactgaactggctcacacattttacattcattgtctcttta	588						
Db	64958	ATTGAGCAGTTGTTGTTCTAGACATTGTTTCCAAACATTTTCCATTTAGCTCATTTA	64899						
Qy	589	agttctagcaattgttaagcagaactatgattattatgactacataaatgaagaataatga	648						
Db	64898	ATCACAAACACCTGAGTAGGTACTACTATTATTTCCCATTTTAAAGATAGGAAACTGA	64839						
Qy	649	gactcagatcacattaaatctcccgagggtcacacagctag	6521						
Db	64838	GGTACTGAACAGTTTAACTACTCTTCTCTGGTTACATGGCTAG	61796						
RESULT 15	AF391809/c								
LOCUS	AF391809	24771 bp	DNA	linear	FBI 20-JUL-2001				
DEFINITION	Homo sapiens coagulation factor II (thrombin) receptor (F2R) gene, complete cds.								
ACCESSION	AF391809								
VERSION	AF391809.2	GI:14971463							
KEYWORDS									
SOURCE	human.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
AUTHORS	1 (bases 1 to 24771) Rieder,M.J., Carrington,D.P., Chung,M.-W., Lee,K.L., Foel,C.L., Yi,O. and Nickerson,D.A.								
TITLE	Direct Submission								
JOURNAL	Submitted (15-JUN-2001) Molecular Biotechnology, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA								
REFERENCE	2 (bases 1 to 24771)								
AUTHORS	Rieder,M.J. and Nickerson,D.A.								
JOURNAL	Submitted (20-JUL-2001) Molecular Biotechnology, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA								
REMARK	Sequence update by submitter								
COMMENT	On Jul 20, 2001 this sequence version replaced oi:14583134. To cite this work please use: SeattleSNPS, NHLBI Program for Genomic Applications, UW-FHCRC, Seattle, WA (URL: http://pga.mbt.washington.edu). Location/Qualifiers								
FEATURES	source	1..24771							
	/organism="Homo sapiens"								
	/db_xref="taxon:9606"								
repeat_region	55..211								
	/rpt_family="MER104"								
variation	234								
	/rpt_type="dispersed"								
	/frequency="0.07"								
	/replace="C"								
variation	352								
	/frequency="0.42"								
	/replace="A"								
variation	536								
	/frequency="0.01"								
	/replace="A"								
variation	709								
	/frequency="0.02"								
	/replace="T"								
repeat_region	811..1058								
	/rpt_family="AluJb"								
variation	1184								
	/rpt_type="dispersed"								
	/frequency="0.01"								
	/replace="G"								
variation	1631..1643								
	/frequency="0.64"								
	/replace="A"								
misc_feature	1545..2063								
	/note="Region not scanned for variation"								

mrna	join(1807..2237,18156..21315) /gene="F2R" /product="coagulation factor II (thrombin) receptor" 1807..21315 /gene="F2R" 2081 /gene="F2R" /frequency="0.01" /replace="A" join(2150..2237,18156..19345) /gene="F2R" /codon_start=1 /product="coagulation factor II (thrombin) receptor" /protein_id="AAK69768.1" /db_xref="GI:14583135" /translation="MGPRLLLVAACFSLGCLPSARTARRPESKATNATILDPRSLF LNPNDKYEPEWDEENESCLTEYRLVLSINKSPLOKOLPAFISDAGCYLTSSWLT LFPSVYTGCVFVSPLNMAIVFVLKMKRPAYVYMLHLATADVLFQSVLFPKIS YFSGSDQFGSELCRFVTAAYFCNMYIASLLMLTISIDRFNAVYPMQSLSRILGR ASFTGLATAIAIAGVPLLDKEQTIOVPLGNITTCHDVNLTELLGTYAYTYSATSA VFVFPFLIISTVCYVSIIRCLSSAVANRKKSRALFSAVFCIFITCEGPIINLLLI AHVSLSHSTITEAAFPAYLLCVCVSSISCCIDPLYYVASSECORYVYSLCKEKS DPSSYNSGOLMASKMDTCSSNLNNSIYKKLT"
variation	2372 /gene="F2R" /frequency="0.05" /replace="A" 2502 /gene="F2R" /frequency="0.01" /replace="A" 2762 /gene="F2R" /frequency="0.06" /replace="T" 2860 /gene="F2R" /frequency="0.40" /replace="A" 2930 /gene="F2R" /frequency="0.22" /replace="C" 2940 /gene="F2R" /frequency="0.01" /replace="G" 2952 /gene="F2R" /frequency="0.01" /replace="T" 3252 /gene="F2R" /frequency="0.01" /replace="C" 3481 /gene="F2R" /frequency="0.39" /replace="C" 3525 /gene="F2R" /frequency="0.10" /replace="I" 3941..4047 /rpt_family="Mir" /rpt_type="dispersed" 3948 /gene="F2R" /frequency="0.21" /replace="G" 4017 /gene="F2R" /frequency="0.24"
repeat_region	
variation	
variation	

```
variation /replace="C"
4038
/ gene="F2R"
/ frequency="0.03"
/ replace="A"
4087
/ gene="F2R"
/ frequency="0.02"
/ replace="T"
4107
/ gene="F2R"
/ frequency="0.03"
/ replace="T"
4336
/ gene="F2R"
/ frequency="0.41"
/ replace="A"
4393
/ gene="F2R"
/ frequency="0.03"
/ replace="A"
4517
/ gene="F2R"
/ frequency="0.07"
/ replace="T"
4925
/ gene="F2R"
/ frequency="0.01"
/ replace="G"
4971..5125
/ rpt_family="MIR"
/ rpt_type=dispersed
5072
/ gene="F2R"
/ frequency="0.01"
/ replace="C"
5128
/ gene="F2R"
/ frequency="0.22"
/ replace="G"
5150..5441
/ rpt_family="AluSx"
/ rpt_type=dispersed
5159
/ gene="F2R"
/ frequency="0.02"
/ replace="T"
5210
/ gene="F2R"
/ frequency="0.01"
/ replace="C"
5276
/ gene="F2R"
/ frequency="0.03"
/ replace="A"
5428
/ gene="F2R"
/ frequency="0.01"
/ replace="A"
5472
/ gene="F2R"
/ frequency="0.40"
/ replace="A"
5480
/ gene="F2R"
/ frequency="0.01"
/ replace="A"
5515
/ gene="F2R"
/ frequency="0.01"
/ replace="A"
```

Query Match

7.0% Score 64.6: DB 9: Length 24771:

```
Best Local Similarity 56.8%: Pred. No. 1.le-05;
Matches 138; Conservative 0; Mismatches 104; Indels 1; Gaps 1:
Oy 556 tgggtcagcattttacattcattgtctctttaagttctagcaatgtgaagcagaacta 615
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9390 TGTGCCAGATACITTCACITTCCTTAGTTCATATATAATTCTTAACATTCTATGACATGAGCA 9331
Oy 616 tgattatgtactacataaataagaagaattgaggctcagatacatttaagtaattctccc 675
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9330 TTATAACTACCATITTTATAATGACAACATGGAGGCACAGAGGCTAAGTAACTAANTCC 9271
Oy 675 agggtcacacagcta-gaactggcagaagcctgggattgacctgatcttccagcattga 734
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9270 AAAGTCACACAGCTATGAAGTAGAGGAGCCAGGGTTTGAACCTAGGTCTGCCAGCTCCA 9211
Oy 735 agaatacataaataataaataactgcaaggccttttctctcagaagaagctcctgcttga 794
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9210 AAGGTCAIACCTTTTGACCACCTCCACGCTATTCTCTCCTTTAAAACTTGTGATTATGAC 9151
Oy 795 cca 797
|||
Db 9150 CTA 9148
```

Search completed: April 29, 2002, 19:50:29
Job time: 22428 sec


```

Db 362 TTCATGCTCTCTGGATTTTCATCAATTAAATNACCT 400
      BE186013
      LOCUS      316 bp mRNA linear EST 22-JUN-2000
      DEFINITION CMI-HT0738-250400-199-b11 HT0738 Homo sapiens cDNA, mRNA sequence.
      ACCESSION BE186013
      VERSION    BE186013.1 GI:8665197
      KEYWORDS  EST.
      SOURCE     human.
      ORGANISM   Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
      1 (bases 1 to 316)
      Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
      Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
      Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
      ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
      Simpson,A.J.
      Shotgun sequencing of the human transcriptome with ORF expressed
      sequence tags
      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
      20202663
      Contact: Simpson A.J.G.
      Laboratory of Cancer Genetics
      Ludwig Institute for Cancer Research
      Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
      Brazil
      Tel: +55-11-2704922
      Fax: +55-11-2707001
      Email: asimpson@ludwig.org.br
      This sequence was derived from the FAPESP/LICR Human Cancer Genome
      Project. This entry can be seen in the following URL
      (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-CMI-HT0738-250
      400-199-b11&t3=2000-04-25&t4=1)
      Seq primer: puc 18 forward
      High quality sequence start: 4
      High quality sequence stop: 316.
      Location/Qualifiers
        1..316
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /clone_lib="HT0738"
          /dev_stage="Adult"
          /note="Organ: Head/neck; Vector: puc18; Site_1: SmaI;
          Site_2: SmaI; A mini-library was made by cloning products
          derived from ORESTES PCR (U.S. Letters Patent application
          No. 196,716 - Ludwig Institute for Cancer Research)
          profiles into the pUC 18 vector. Reverse transcription of
          tissue mRNA and cDNA amplification were performed under
          low stringency conditions."
          BASE COUNT 69 a 93 c 71 g 82 t 1 others
      ORIGIN
        1
      Query Match 61.2%; Score 293.6; DB 9; Length 316;
      Best Local Similarity 98.4%; Pred. No. 6.6e-40;
      Matches 307; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
      1:
      QY 60 agcagtcctgtagtcttggagatttccatcttctgggtctc-tgcccgaaatccgacaa 118
      DB 5 AGCAGCTCTGTTACTCTGGGAGTTCCATCTTTCCGGTCTCGTATCAGATCCGACAA 64
      QY 119 cagctgtccagctgacacgtaccagctactggtctgctgatgatgaagccctgagt 178
      DB 65 CAGCTGCTCAGCTGACAGCTATCCAGCTACGGGTCTGCTGATGATGAGAGCCCTGATG 124
      QY 179 ctgaaccactgctgtcgaaccactgcaccactgctgctctac-aactgcaccaccg 238
      DB 125 CTGNAACCACTGCTGCTGCAACCACTGCGACCACTGCTGCTCTCTACCACTGCAACCCG 184

```

```

QY 239 ctgcttctaccactgctgttaagacattccagttttaccacaaatgggttggaatctcc 298
      DB 185 CTGCTTCTACCACCTGCTGTAAGACATTCAGTTTACCCAAATGGGTGGGATCTCC 244
      QY 299 cgaatggttagagtggtgctccctgaatgaatcagcttgaagctctctcgaattggtcaca 358
      DB 245 CGAATGGTAGAGTGTGCTCCCTGAGATGAATCAGCTTGAGTCTTCTGCAATGGTCACAA 304
      QY 359 ctattcatgctt 370
      DB 305 CTATTCATGCTT 316
      RESULT 15
      BE925938
      LOCUS      316 bp mRNA linear EST 02-OCT-2000
      DEFINITION CMI-BT0738-240800-199-b11 BT0738 Homo sapiens cDNA, mRNA sequence.
      ACCESSION BE925938
      VERSION    BE925938.1 GI:10453014
      KEYWORDS  EST.
      SOURCE     human.
      ORGANISM   Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
      1 (bases 1 to 316)
      Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
      Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
      Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
      Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
      ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
      Simpson,A.J.
      Shotgun sequencing of the human transcriptome with ORF expressed
      sequence tags
      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
      20202663
      Contact: Simpson A.J.G.
      Laboratory of Cancer Genetics
      Ludwig Institute for Cancer Research
      Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
      Brazil
      Tel: +55-11-2704922
      Fax: +55-11-2707001
      Email: asimpson@ludwig.org.br
      This sequence was derived from the FAPESP/LICR Human Cancer Genome
      Project. This entry can be seen in the following URL
      (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-CMI-BT0738-240
      800-199-b11&t3=2000-08-24&t4=1)
      Seq primer: puc 18 forward
      High quality sequence start: 4
      High quality sequence stop: 316.
      Location/Qualifiers
        1..316
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /clone_lib="BT0738"
          /dev_stage="Adult"
          /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
          SmaI; A mini-library was made by cloning products derived
          from ORESTES PCR (U.S. Letters Patent application No. 196
          716 - Ludwig Institute for Cancer Research) profiles
          into the pUC 18 vector. Reverse transcription of tissue
          mRNA and cDNA amplification were performed under low
          stringency conditions."
          BASE COUNT 69 a 93 c 71 g 82 t 1 others
      ORIGIN
        1
      Query Match 61.2%; Score 293.6; DB 10; Length 316;
      Best Local Similarity 98.4%; Pred. No. 6.6e-40;
      Matches 307; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
      1:
      QY 50 agcagtcctgtagtcttggagatttccatcttctgggtctc-tgcccgaaatccgacaa 118

```

```

.M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, I., Soares, M., Tan, F., Trevaskis, E., Waterston
R., Williamson, A., Wohldmann, P., and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 516
High quality sequence stops: 307
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 516      Std Error: 0.00
Seq primer: M13Rp1
High quality sequence stop: 307.
      Location/Qualifiers
          1. 403
              /organism="Homo sapiens"
              /db_xref="CDB:572980"
              /db_xref="taxon:9606"
              /clone="IMAGE:158921"
              /clone_lib="Soares breast 2NBHBst"
              /sex="Female"
FEATURES             source

```

```

/adb_host= DHI08 (ampicillin resistant)
/Note=Organ: breast; Vector: p7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; lnt
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGGCGGCCCTTTTITTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of a modified p7T3 vector (Pharmacia).
Library went through one round of normalization to a Cot
230. Library constructed by Bento Soares and M.Fatima
Bonaldo.

```

BASE COUNT		82 a	113 c	89 g	116 t	3 others
ORIGIN						
Query Match 62.9%; Score 302; DB 10; Length 403;						
Best local Similarity 93.0%; Pred. No. 2.4e-41;						
Matches 371; Conservative 0; Mismatches 18; Indels 10; Gaps 5						
Oy	10	gcattttgtctgtccctgattcttcaggtcaccacattgaagttcttagcagctctg	69			
Db	2	GCAITTTTGTCTGCTCCCTGATCTTCAGGTCAACCACCAATGAAGTCTTTAGCAGTCTG	61			
Oy	70	gtactctggaagttccattcttctgtctctccagaatccgaacagctgtctcca	129			
Db	52	GNACICTTGGGAGTTTCCATCTTCTGTGCTCTGCCAGAAATCCGACACAGCTGTCTCCA	121			
Oy	130	gctgacagatctccagctactgctgtctgctgatgaagccctctgatgctgaaccact	189			
Db	122	GCTGACACGTATCCAGCTACTTGCTCTCTGATGATGAAGCCCTGTGCTGAACCACT	181			
Oy	190	gctctgcaaacactgcaacactgactctcctaccactgcaacacgctgtcttacc	249			
Db	182	GCTGCTGCACCACTTGCACCACTGCTGCTCTCTACCCTGCACCAACCGCTGCTTCTACC	241			
Oy	250	actgtcgtaaagacatcttcaggttttacccaaatgggtgggagatctcccgaaatgg--ta	307			
Db	242	ACTGTCTGTAAAGACATTCACGTTTTACCCAAATGGGTGGGGATCTCCCGAATGGGTAG	301			
Oy	308	aaagtgtacctgaag--atgaatcagcttaagctct---tctgcaattggtcacaacta-	361			
Db	302	GAGTGTGTCCTTCCAGATGGGAATCAGCTTGAGTCTTTCTGGCAATTGGGTTCACAANTAT	361			
Oy	352	ttcatgcttctctgltg--atttcatccaaactacttacct 398				

```

FEATURES
  source          Location/Qualifiers
    1..343
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_lib="Athensys RAGE Library"
      /cell_line="HT1080"
      /note="See 'Creation of Genome-wide protein Expression
      Libraries Using Random Activation of Gene Expression',
      Nature Biotechnology, in press. Note that even though the
      cell type indicated is HT1080, since a random activation
      method was used, these sequence tags are not necessarily
      expressed in HT1080 under normal circumstances."
    99 a          53 c          57 g          80 t          4 Others
BASE COUNT
ORIGIN
Query Match          65.0%; Score 311.8; DB 10; Length 343;
Best Local Similarity 93.9%; Pred. No. 5.7e-43;
Matches 322; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
Qy 107 agaatccaacagcgtctccagctgacacgtatccagctactggtcctcgtgatgatg 166
      |||||||
Db 343 AGAAATCTGCAACAGCTGCTCCAGCTGACACGATATCCAGCTACTGGCTCTGCTGAATG 284
      |||||||

```

[illegible]

REFERENCE
AUTHORS
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Regai, M.A., da Silva, W. Jr., Zaia, M.A., Bordin, S., Costa, F.F.,
Goldman, G.B., Carvalho, A.F., Matsumura, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., Carvalho, P., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000).
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922


```

Db 133 GCTGCTCCAGCTGACACGATCCAGCTACTGGTCTCTGCTGATGAAGCCCTGATGTT 192
Oy 181 gaaccactgtgtgaaaccactgacacactgtctctctaccactgcaaccacgct 240
Db 193 GAAACACCTGTGTCGAACACCTGCGACCACTGCTGCTCTCCACCTGCAACCCGCT 252
Oy 241 gctctaccactgtctgaaagacattccagttttaccacaaatgggtggggtctcccg 300
Db 253 GCTTCTACCACTGCTGTAAGACATTCAGTTTTACCAAAATGGGTGGGATCTCCCG 312
Oy 301 aatgataagtgatctccctgagatggaatcagcttgagttctctcgaattggtcacaact 360
Db 313 AATGGTAGAGTGTGCTCCGATGAATGGAATCAGCTTGAGTCTTCGCAATGGGTCACACT 372
Oy 361 attcagctctctgagattccacactaccacttacccttgctcctacagata 408
Db 373 ATTATGCTTCTGATTTTCAATCCAACTACTTACCTTGCCTACACTA 420

```

```

RESULT 5
AI139456/c 402 bp mRNA linear EST 27-OCT-1998
LOCUS qc20g01.x1 Soares fetal heart NbHL19W Homo sapiens cDNA clone
DEFINITION IMAGE:1710192 3' similar to contains element MER40 repetitive
element ;, mRNA sequence.
ACCESSION AI139456
VERSION AI139455.1 GI:3645428
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 402)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rcp@pshs.fda.gov
This clone is available royalty-free through LLNL: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 510 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 398.
Location/Qualifiers
1. .402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1710192"
/clone_lib="Soares_fetal_heart_NbHL19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGCGGCGGCACTTTTITTTTTTTTTT 3'].
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot - 5. Library constructed by:
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHL19W."

```

```

BASE COUNT 115 a 72 c 115 g 100 t
ORIGIN

Query Match 92.84; Score 397.2; DB 9; Length 402;
Best Local Similarity 99.34; Pred. No. 2.7e-57;
Matches 399; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Oy 67 clggactcttggagattlccatcttcttggtctctgcccagaatccgacacagctgct 126
Db 402 CTGGTACTCTTGGGAGTTTCCATCTATCTGCTCTGCCCCAAGATCCGACACAGCTGCT 343
Oy 127 ccagctgacacgtatccagctactgctgctgctgctgctgctgctgctgctgctgct 186
Db 342 CCAGCTGACACGTATCCAGCTACTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 283
Oy 187 actgctgctgcaacacctcgcgacacactgctgctctctaccactgcaacacgctgctct 246
Db 282 ACTGCTGCTGCAACCACTGCGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 223
Oy 247 accactgctctgaaagacattccagttttaccacaaatgggtggggtcctccgaatggt 306
Db 222 ACCACTGCTGCTGTAAGACATTCAGTTTTACCAAAATGGTGGGATCTCCCGAATGCT 153
Oy 307 agagtgtgctccctgagatggaatcagctgagctctctcgaattggtcacaactattcat 366
Db 152 AGAGTGTGCTGAGATGGAATCAGCTTGAGTCTTCTGCAATGGTGCACAACTATTCAT 103
Oy 367 gcttctgctgattctccacactaccacttaccctgctcctacgataccccctttatctc 426
Db 102 GCCTTCTGCTGATTTTCAATCCAACTACTTACCTTGCCTACGATATCCCTTTATCT 43
Oy 427 agtttatcttcttcaataaaataaataaataaataaataaataaataaataaataa 468
Db 42 AGTTTATTTCTTTTCAATAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1

RESULT 5
BG202312 464 bp mRNA linear EST 21-APR-2001
LOCUS RST121568 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION BG202312
ACCESSION BG202312
VERSION BG202312.1 GI:13723999
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 464)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittinton,J.,
Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith
,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
,J., Danzig,J. and Ducar,M.
Creation of genome-wide protein expression libraries using random
activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
21227151
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 464.
Location/Qualifiers
1. .464
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/clone_line="HT1080"
/notes="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."

```

```

BASE COUNT 126 a 119 c 91 g 128 t
ORIGIN

```


Tue Apr 30 14:17:50 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2002, 17:16:35 ; Search time 6783.7 seconds
(without alignments)
955.015 Million cell updates/sec

Title: US-09-248-178-56
Perfect score: 480
Sequence: 1 ggccttgagcattttgtc.....gcaacaaaaaaaaaaaaa 480

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues 27472414

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estnu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_estl.*
10: gb_est2.*
11: gb_hic.*
12: gb_gss.*
13: em_gss_hum.*
14: em_gss_inv.*
15: em_gss_pln.*
16: em_gss_vrt.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	422.4	88.0	436	10 W72837	W72837 zd57g12.r1
C 2	418.4	87.2	420	9 AW070916	AW070916 xat1d01.x
C 3	418.4	87.2	432	10 W72838	W72838 zd57g12.s1
C 4	403.2	84.0	423	10 BF742809	BF742809 I12-BT080
C 5	397.2	82.8	402	9 A1139456	A1139456 qc20g01.x
6	372.8	77.7	464	10 BG202312	BG202312 RST21668
7	358.6	74.7	376	10 BG218084	BG218084 RST37810
8	337.4	70.3	391	10 BG185269	BG185269 RST4208 A
9	314	65.4	314	12 AG594036	AG594036 AST-2H8BG
C 10	311.8	65.0	343	10 BG202313	BG202313 RST21669
C 11	306.6	63.9	345	9 BE092421	BE092421 CM1-BT073
C 12	302	62.9	403	10 R75793	R75793 y121f09.r1
13	293.6	61.2	316	9 BE186013	BE186013 CM1-BT073
14	293.6	61.2	316	10 BE926938	BE926938 CM1-BT073
15	293.6	60.6	294	9 AA340069	AA340069 EST45219
16	291	60.6	294	9 AA340069	AA340069 EST45219
C 17	259	54.0	443	9 A1127172	A1127172 qb99a08.x

18	251.2	52.3	308	9	A1905687	A1905687 IL-BT095-
C 19	224.8	46.8	462	9	AW294149	AW294149 UI-H-B12-
C 20	224	46.7	224	9	AW291950	AW291950 UI-H-B12-
21	219.4	45.7	251	10	BG197186	BG197186 RST16424
C 22	204.4	42.6	221	9	BE074512	BE074512 IL5-BT057
23	200.2	41.7	230	10	BE815819	BE815819 PM3-BN017
24	189.4	39.5	248	10	BE815824	BE815824 PM3-BN017
25	177	36.9	211	10	BG183176	BG183176 RST2047 A
26	176	36.7	212	10	BG192597	BG192597 RST11712
27	165.4	34.7	211	10	BG207535	BG207535 RST27017
28	161	33.5	193	9	A1905624	A1905624 CM-BT094-
C 29	156	32.5	224	9	A1905837	A1905837 IL-BT099-
C 30	150.4	31.3	178	9	A1905623	A1905623 CM-BT094-
C 31	133	27.7	136	9	A1905633	A1905633 CM-BT094-
C 32	106.6	22.2	498	10	B1279448	B1279448 UI-R-DA0-
C 33	105.8	22.0	492	10	B1279431	B1279431 UI-R-DA0-
C 34	103.8	21.6	224	10	BE815853	BE815853 PM3-BN017
C 35	103.4	21.5	486	10	B1282803	B1282803 UI-R-DA0-
C 36	103.4	21.5	490	10	B1279159	B1279159 UI-R-DA0-
C 37	103.4	21.5	490	10	B1279279	B1279279 UI-R-DA0-
C 38	103.4	21.5	493	10	B1279513	B1279513 UI-R-DA0-
C 39	103.4	21.5	493	10	B1279518	B1279518 UI-R-DA0-
C 40	103.4	21.5	495	10	B1279214	B1279214 UI-R-DA0-
C 41	103.4	21.5	497	10	B1279243	B1279243 UI-R-DA0-
C 42	103	21.5	487	10	B1279415	B1279415 UI-R-DA0-
C 43	101.8	21.2	499	10	B1279473	B1279473 UI-R-DA0-
C 44	101.8	21.2	500	10	B1279161	B1279161 UI-R-DA0-
C 45	101.8	21.2	503	10	B1279551	B1279551 UI-R-DA0-

ALIGNMENTS

RESULT 1
W72837 436 bp mRNA linear EST 16-OCT-1996
LOCUS zd57g12.r1 Soares_fetal_heart NBHH19W Homo sapiens cDNA clone
DEFINITION IMAGE:344806 5' similar to contains element MER40 repetitive
element ;, mRNA sequence.

ACCESSION W72837
VERSION W72837.1 GI:1382813

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 436)

AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Riffin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE The WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL: contact the
Insert Consortium (info@image.llnl.gov) for further information.
Insert Length: 508 Std Error: 0.00
Seq primer: mob.REGA.ET
Location/Qualifiers
1. 436
/organism="Homo sapiens"
/db_xref="GDB:1270181"
/db_xref="taxon:9606"
/clone="IMAGE:344806"
/clone_lib="Soares_fetal_heart_NBHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"


```

CDS
complement(join(6102..6251,6309..6712,6961..7000))
/gene="tra"
/note="tra gene product; Nucleotide sequence of the Celera
sequence differs from the published sequence for this
transcript."
/codon_start=1
/db_xref="FLYBASE:FBan0016724"
/db_xref="FLYBASE:FBgn0003741"
/protein_id="AAF49441.1"
/db_xref="GI:7294086"
/translation="MKWDASSGTOHRSRSGSRSPRRPEYHGRSEPDSPRKKHEKI
PIFADEQDRLRLRQRRAHOSRTRRSRSCSSIRSRHRRHQRSPRRNRBSR
SSERKRRHRRHRRHRRHRRHRRHRRHRRHRRHRRHRRHRRHRRHRRHRRH
RLRPFPFPAPYRQRPPEIGVREFRYNAGPPY"
complement(join(7163..9959,10021..10777,10839..11000))
/mRNA
/gene="BcDNA:LD24702"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/product="CI38259"
/db_xref="FLYBASE:FBan0017286"
/db_xref="FLYBASE:FBgn0027500"
complement(<7163..>11000)
/gene="BcDNA:LD24702"
/note="CG17286"
/map="73A7-73A9"
/db_xref="FLYBASE:FBan0017286"
/db_xref="FLYBASE:FBgn0027500"
complement(join(7361..9959,10021..10777,10839..10923))
/gene="BcDNA:LD24702"
/note="BcDNA:LD24702 gene product; Nucleotide sequence of
the Celera sequence differs from the published sequence
for this transcript"
/codon_start=1
/db_xref="FLYBASE:FBan0017286"
/db_xref="FLYBASE:FBgn0027500"
/protein_id="AAF49440.2"
/db_xref="GI:10727879"
/translation="MDSSSGSGNGSEMDONSLGILNMDNLKVFGLSSFSKSGRRSGE
TGSDPAVHERLEERLRALETLEKPPRSOAKSTALOMRLSTNISELVITDITLNL
NESSFLVPATKNTGISFPAEILIGSTLCQAGKORREKPSLVAEILKLSSEFVEK
ALLERKLVQAOASVHSLRGSSSLDSFNCRSRLPRTADMSELNLGGGGSFGF
SASLAAGODESFPAELQKSLVGEISWAOEFTAMPRTSLTQALKNPILDSVSGSE
TSVNSVLAGDDFDGLNYFOTRSNIWNIVSNPSNTQPLLEPESNITLDSVSGSE
ETLPQDRIYTKTDAITGSLNLRMLKMOODELETALKSRNGLAAKETKRPPSSSE
ILSLAIDKALPDIDLNSTSTVEVNLHMGHNGNDDGENEMOSSNSHAERTSGE
GSKNLTDMSTFDSVLNDFRHLQOSTSRKPLSLADHPQITISRADTDPVTEACA
DIDWSPVKEPSRRYRTRIKLSRAASPSSDGVRPLICTEDENDEDEEDKTPVKK
RVSITMGLPIKPSRSLPDLGCDVAASSTEDFGISPNLAKNLSPLSPPSCL
SPLLDSTSDRLRMGAARKNARNSPAGSEASSTGFTASGRGLGHTSSVAVPRISL
PCSNSSEFSHRDQKLLPLPHTHTILCWGSKTLEIDVRKSMQVKTADRLVIRLGIOG
Query Match 12.9%; Score 77.6; DB 3: Length 280897;
Best Local Similarity 54.1%; Pred. No. 9,1e-08;
Matches 158; Conservative 0; Mismatches 134; Indels 0; Gaps 0
Qy 310 cccgcgtggctacaccacgcgcctcgagatcgtagggcaacatccggaagcagaag 369
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280218 CTCGCAAGGAGTACACCGCAGGATACAAATTTCATCGGACACATACCGCAACGACG 280277
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 370 aagagatcaccaagatctctctgatacgaagagcttcagaagaaatacaactccctat 429
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280278 CGGATATCTACAAATGTCGTGATGACACGGACACACTCCAAAGACGACTTAATGTCGTG 280337
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 430 ctgggaagctgacacgaacgtttgcgtgactgatgactgttgttcaagatgccaaag 489
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280338 GTGCCACGCTTCAGCGCGAGTTCACACTACACGGACGACCTGCTCTTCACAGCGCCCAAGC 280397
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 490 agacagatgctgttcgaagcctataagtatctagctgctgcacgaacactgcagcc 549
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280398 ATGATCTCCACGGAAGCGGCGCTACCAAGTACTTGGCCCAACTGCACGCCAATTCGCAATG 280457
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 550 agctcatcacgaacatcgagacacagaccatcatcgagaggtttcagag 601
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 280458 AGCTGTCGAGTGC GTCTCCCTCACGGGNAACGTGACCANGCAGATCCCGCA 280509

Search completed: April 29, 2002, 19:38:14
Job time: 21693 sec

```

VERSION      AE003526.2  GI:10727878
KEYWORDS     HTG
SOURCE       fruit fly:
ORGANISM     Drosophila melanogaster
REFERENCE     Fukuyota; Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:
AUTHORS       Pterygota; Neoptera; Endopterygota; Diptera: Brachycera;
              Muscomorpha; Ephydroidea: Drosophilidae; Drosophila.
              1 (bases 1 to 280887)
              Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,
              Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,
              George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,
              Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,
              Brannon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D.,
              Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor
              Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J.,
              Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A.,
              Baxter, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y.,
              Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkov, D.,
              Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Butts, K.C.,
              Puzan, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I.,
              Cherry, J.M., Cawley, S., Dahlke, C., Davenport, I.B., Davies, P., de
              Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M.,
              Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C.,
              Dunn, P., Durbin, K.J., Evans, M., Ferraz, C., Ferreira, S.,
              Fleischmann, W., Fosler, C., Gabriellian, A.E., Garg, N.S.,
              Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z.,
              Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J.,
              Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,
              Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,
              Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
              Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A.,
              Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattel, B., McIntosh, T.C.,
              McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Moharry, C.,
              Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,
              Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,
              Nusskern, D.R., Pacle, J.M., Palazzolo, M., Pittman, G.S., Pan, S.,
              Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K.,
              Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I.,
              Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C.,
              Stapleton, M., Strong, R., Sun, E., Svirskaas, R., Tector, C., Turner, R.,
              Venier, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A.,
              Weinstein, G.M., Weissbach, J., Williams, S.M., Woodage, T.,
              Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Ye, R.F.,
              Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H.,
              Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O.,
              Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
              The genome sequence of Drosophila melanogaster
              Science 287 (5461), 2185-2195 (2000)
              20196006
              2 (bases 1 to 280887)
              Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, J.C.
              Direct Submission
              Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
              Rockville, MD, USA
              On Oct 9, 2000 this sequence version replaced gi:7294043.
              Location/Qualifiers
              1..280887
              /organism="Drosophila melanogaster"
              /db_xref="taxon:7227"
              /chromosome="3L"
              join(81..401,546..679,736..1039,1142..2073)
              /gene="CG4169"
              /product="CT13760"
              /db_xref="FLYBASE:FBan0004169"
              /db_xref="FLYBASE:FBgn0035642"
              <81..>2073
              /gene="CG4169"
              /map="73A7-73A9"
              /db_xref="FLYBASE:FBan0004169"
              /db_xref="FLYBASE:FBgn0035642"
              join(357..401,546..679,736..1039,1142..1981)
              /gene="CG4169"
              /note="CG4169 gene product"

/codon_start=1
/db_xref="FLYBASE:FBan0004169"
/db_xref="FLYBASE:FBgn0036642"
/protein_id="AAF4944.1"
/db_xref="GI:7294089"
/translation="MACNASKTSLRAIAKRGVATCPRPVGDLSAVNVKVLKLVLA
TADATLPVSRLVGLGAGSESYDIQASHLLRLAGGLSTONSTAFALARNIOOVGG
EERAIELVHRKAFNGLNHSYSPFOLGKLSLSELSLHVAQITFAAGRAVYGVGLDN
NTILAGFAOTLQFPSSGSKAANNYGGDKDKTSGHRAVAVAGOGAASNHKEALF
ATLEQALAKAATKRGTSAGLGEAYNCAGGVGASVAVNASTSDAGLFGVVSADSK
DICKTVEFLVRLKASVSDKDVARGKALLKARIISRYSSDGLIKRIGROALTRNV
LEADALGATDGISQSQOQEAANKVGSKLAVGAIGHANVPYASDLA"
complement(join(2184..2344,2398..>2991))
/gene="CG4109"
/product="CT13642"
/db_xref="FLYBASE:FBan0004109"
/db_xref="FLYBASE:FBgn0036643"
complement(<2184..>2991)
/gene="CG4109"
/map="73A7-73A9"
/db_xref="FLYBASE:FBan0004109"
/db_xref="FLYBASE:FBgn0036643"
complement(join(2184..2344,2398..2935))
/gene="CG4109"
/note="CG4109 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0004109"
/db_xref="FLYBASE:FBgn0036643"
/protein_id="AAF4944.1"
/db_xref="GI:7294088"
/translation="MALVDHSDWIEYEGGERLHQLLVLNORQNLNPTRSQFVQLI
SSRTIGIQALAKMHLKVLVDNDAITWETSPPEELQORRIDWDLTISOLREKFKAN
SSRSNVPAAGSAMDODLQPGHSNRTALDVEALKOKKTEMAQOENEGLEVLST
LSROPLATOLGNEVEDONNILDNLANAMDRVETGVQRETOSIGOVNRDSTWGYLV
IHALFVALIVVVFV"
join(<4809..4914,4977..5129,5204..5487,5606..>5731)
/gene="CT1373Ah"
/product="CT13794"
/db_xref="FLYBASE:FBan0004195"
/db_xref="FLYBASE:FBgn0002283"
<4809..>5731
/gene="CT1373Ah"
/note="CG4195"
/map="73A7-73A9"
/db_xref="FLYBASE:FBan0004195"
/db_xref="FLYBASE:FBgn0002283"
join(4809..4914,4977..5129,5204..5487,5606..5731)
/gene="CT1373Ah"
/note="CT1373Ah gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0004195"
/db_xref="FLYBASE:FBgn0002283"
/protein_id="AAF4944.1"
/db_xref="GI:7294087"
/translation="MERRVKIKTIINPHITCKICGGYFIDATITVTECLHTFCKSLVKH
LEEKKTCTPCDNIHHOSHPLOYISDFRTMODIVYKLVKQDESRERDFYKSRNMP
CPKDIIONHEDNEKVMADAFSDHRLDEOVNVCLEISNNFKNLQRRFIRCSQAT
ITHLXLVAKKILNGIEKYREIDILCNELLGKDHITLKEVYVTRWRFRDPLRLQFRP
RVEL"
complement(join(<6102..6251,6309..6712,6961..>7000))
/gene="tra"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/product="CT37211"
/db_xref="FLYBASE:FBan0016724"
/db_xref="FLYBASE:FBgn0003741"
complement(<6102..>7000)
/gene="tra"
/note="CG16724"
/map="73A7-73A9"
/db_xref="FLYBASE:FBan0016724"
/db_xref="FLYBASE:FBgn0003741"

```


the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base

```

Query Match      12.9%; Score 77.6; DB 3: Length 181437;
Best Local Similarity 54.1%; Pred. No. 9.1e-08;
Matches 158; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

Oy 310 cccggtgctccaccagcgactcttgagatcggtggcaacatccgggaagcagaag 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 53171 CTCGCAAGGAGTACACCCGAGGATACATGAATTCATCGGGAACATACGCAAGCAAG 53112

Oy 370 aagagatcccaagatctgtctgatacgaaggagccttcaagaagaatcaactccctat 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 53111 CGGATATCTACAAAGTTCTGGATGACACGGCAGCAACTCCAAAGCAGCTTAATGTCGTGG 53052

Oy 430 ctgggaagctggaacggacgtttgcgtgactgatgactgtgttcaaggatgccaa 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 53051 GTGCCAGTTCGAGCGCCAGTTCACTACACGGAGCAGCTGCTCTCCAGAGCGCAAGC 52992

Oy 490 aggaagatctgtcgaagcctataagatctagctgtctgcacgagaaactcagcc 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 52991 ATGATCTCCACGGAAGCGGGCCIAAAGCTACTGCCCCCACTGACGCCCAATGCAATG 52932

Oy 550 agctatccagaccatcgaggacacagcaccatcatcgaggagatttcgaga 601
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 52931 AGCTGTCGAGTCTGCTCCTCCAGGGAACGTGACCAAGCAGATCCGCGA 52880
  
```

```

RESULT 14
AC006933/c 186002 bp DNA linear INV 09-JUL-1999
LOCUS Drosophila melanogaster, chromosome 3L, region 73B7-73D5, BAC clone
DEFINITION PACR48E21, complete sequence.
ACCESSION AC006933
VERSION AC006933.3 GI:5042394
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 186002)
Celinker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhof,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,P.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Ling,H.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (02-MAR-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Jun 11, 1999 this sequence version replaced gi:4836865.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgpf@fruitfly.berkeley.edu.
Location/Qualifiers
    1..186002
    /organism="Drosophila melanogaster"
    /strain="y; cn bw sp"
    /db_xref="taxon:7227"
    /chromosome="3L"
    /map="73B7-73D5"
    /clone="BAC PACR48E21 (D458)"
    /clone_lib="RPC1-98 (Roswell Park Cancer Institute
    Drosophila melanogaster BAC library, partial EcoRI in
    pBACe3.6)"
    pBACe3.6
BASE COUNT 54229 a 38636 c 39135 g 54002 t
ORIGIN
  
```

Pfeiffer,B., Poon,L., Richards,S., Sequeira,A., Sethi,H.,
Svirskas,R.R., Wan,K.H., Webster,D., Woolley,P., Yang,S., Yee,M.,
Zieran,L.L. and Rubin,G.M.
Sequencing of Drosophila chromosome 3L, region 73B7-73D5
Unpublished (1998)
2 (bases 1 to 186002)
Celinker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhof,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,P.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (02-MAR-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Jun 11, 1999 this sequence version replaced gi:4836865.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgpf@fruitfly.berkeley.edu.

FEATURES

source

1..186002

/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3L"
/map="73B7-73D5"
/clone="BAC PACR48E21 (D458)"
/clone_lib="RPC1-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACe3.6)"

BASE COUNT 54229 a 38636 c 39135 g 54002 t

ORIGIN

```

Query Match      12.9%; Score 77.6; DB 3: Length 186002;
Best Local Similarity 54.1%; Pred. No. 9.1e-08;
Matches 158; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

Oy 310 cccggtgctccaccagcgactcttgagatcggtggcaacatccgggaagcagaag 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35563 CTCGCAAGGAGTACACCCGAGGATACATGAATTCATCGGGAACATACGCAAGCAAG 35504

Oy 370 aagagatcccaagatctgtctgatacgaaggagccttcaagaagaatcaactccctat 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35603 CGGATATCTACAAAGTTCTGGATGACACGGCAGCAACTCCAAAGCAGCTTAATGTCGTGG 35544

Oy 430 ctgggaagctggaacggacgtttgcgtgactgatgactgtgttcaaggatgccaa 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35543 GTGCCAGTTCGAGCGCCAGTTCACTACACGGAGCAGCTGCTCTCCAGAGCGCAAGC 35484

Oy 490 aggaagatctgtcgaagcctataagatctagctgtctgcacgagaaactcagcc 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35483 ATGATCTCCACGGAAGCGGGCCCTACAAAGCTACTGCCCCCACTGACGCCCAATGCAATG 35424

Oy 550 agctatccagaccatcgaggacacagcaccatcatcgaggagatttcgaga 601
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35423 AGCTGTCGAGTCTGCTCCTCCAGGGAACGTGACCAAGCAGATCCGCGA 35372
  
```

RESULT 15

AE003526
LOCUS Drosophila melanogaster genomic scaffold 142000013386050 section 42
DEFINITION of 54 complete sequence.
ACCESSION AE003526 AE002602

AE003526
Drosophila melanogaster genomic scaffold 142000013386050 section 42
of 54 complete sequence.
AE003526 AE002602

AE003526
Drosophila melanogaster genomic scaffold 142000013386050 section 42
of 54 complete sequence.
AE003526 AE002602

- JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
- Unpublished
2 (bases 1 to 181437)
Worley, K.C.
Direct Submission
Submitted (22-SEP-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 181437)
- REFERENCE
AUTHORS
- Worley, K.C., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alstbrooks, S.L., Amaratunga, H.C., Are, J.R., Banks, T., Barbarta, J.,
Benton, J., Bimaye, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
Burck, P., Burckett, C., Burrell, K.L., Byrd, N.C., Cartron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Drepper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,
Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Louis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Louisseg, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Maxhiney, E., McLeod, M.P., Meador, M.,
Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nkwenkwo, S.,
Ogih, M., Okuwonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,
Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N.,
Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wlezyk, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
- TITLE
JOURNAL
REFERENCE
AUTHORS
- Direct Submission
Submitted (10-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 181437)
- REFERENCE
AUTHORS
- Worley, K.C., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alstbrooks, S.L., Amaratunga, H.C., Are, J.R., Banks, T., Barbarta, J.,
Benton, J., Bimaye, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
Burck, P., Burckett, C., Burrell, K.L., Byrd, N.C., Cartron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Drepper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,
Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Louis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Louisseg, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Maxhiney, E., McLeod, M.P., Meador, M.,
Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nkwenkwo, S.,
Ogih, M., Okuwonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,
Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N.,
Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wlezyk, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
- TITLE
JOURNAL
COMMENT
- Direct Submission
Submitted (03-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 10, 2001 this sequence version replaced gi:8101327.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu
- CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of

*	14633	18437	contig of 3805 bp in length
*	14638	18537	gap of unknown length
*	18538	22123	contig of 3586 bp in length
*	22124	22223	gap of unknown length
*	22224	25731	contig of 3508 bp in length
*	25732	25831	gap of unknown length
*	25832	28890	contig of 3059 bp in length
*	28891	28990	gap of unknown length
*	28991	31876	contig of 2886 bp in length
*	31877	31977	gap of unknown length
*	31977	34726	contig of 2750 bp in length
*	34727	34826	gap of unknown length
*	34827	37144	contig of 2318 bp in length
*	37145	37244	gap of unknown length
*	37245	40483	contig of 3139 bp in length
*	40384	40483	gap of unknown length
*	40484	43322	contig of 2839 bp in length
*	43323	43322	gap of unknown length
*	43323	45399	contig of 1977 bp in length
*	43423	45499	gap of unknown length
*	45400	47906	contig of 2407 bp in length
*	47907	48009	gap of unknown length
*	48007	50859	contig of 2853 bp in length
*	50860	50859	gap of unknown length
*	50960	54078	contig of 3119 bp in length
*	54079	54178	gap of unknown length
*	54179	56138	contig of 1960 bp in length
*	56139	56238	gap of unknown length
*	56239	59204	contig of 2966 bp in length
*	59205	62250	gap of unknown length
*	59305	62250	contig of 2946 bp in length
*	62251	62350	gap of unknown length
*	62351	65162	contig of 2812 bp in length
*	65163	65262	gap of unknown length
*	65263	67095	contig of 1833 bp in length
*	67096	67195	gap of unknown length
*	67196	69359	contig of 2164 bp in length
*	69360	69459	gap of unknown length
*	72018	72018	contig of 2559 bp in length
*	72019	72118	gap of unknown length
*	72119	74575	contig of 2457 bp in length
*	74576	74675	gap of unknown length
*	74676	76752	contig of 2077 bp in length
*	76753	76852	gap of unknown length
*	76853	78542	contig of 1630 bp in length
*	78543	78642	gap of unknown length
*	78643	79814	contig of 1172 bp in length
*	79815	79914	gap of unknown length
*	79915	81212	contig of 1288 bp in length
*	81213	81312	gap of unknown length
*	81313	84320	contig of 3008 bp in length
*	84321	84420	gap of unknown length
*	84421	86815	contig of 2335 bp in length
*	86816	90121	contig of 3206 bp in length
*	90122	90221	gap of unknown length
*	90222	92344	contig of 2113 bp in length
*	92345	92434	gap of unknown length
*	92435	94033	contig of 1599 bp in length
*	94034	94133	gap of unknown length
*	94134	96100	contig of 1967 bp in length
*	96101	96200	gap of unknown length
*	96201	97631	contig of 1431 bp in length
*	97632	97731	gap of unknown length
*	97732	99389	contig of 1658 bp in length
*	99390	99489	gap of unknown length
*	99490	101608	contig of 2119 bp in length
*	101609	103260	contig of 1552 bp in length
*	103261	103360	gap of unknown length
*	103361	105942	contig of 2582 bp in length
*	105943	107862	gap of unknown length
*	106043	107786	contig of 1744 bp in length

	* 107767	107886:	gap of unknown length
	* 107867	109676:	contig of 1792 bp in length
	* 105679	109778:	gap of unknown length
	* 109779	111189:	contig of 1411 bp in length
	* 111150	111289:	gap of unknown length
	* 111290	112591:	contig of 1302 bp in length
	* 112592	112691:	gap of unknown length
	* 112692	113758:	contig of 1067 bp in length
	* 113759	113858:	gap of unknown length
	* 113859	115175:	contig of 1321 bp in length
	* 115180	115275:	gap of unknown length
	* 115280	116458:	contig of 1219 bp in length
	* 116459	116558:	gap of unknown length
	* 116559	117774:	contig of 1176 bp in length
	* 117775	117874:	gap of unknown length
	* 117875	119245:	contig of 1371 bp in length
	* 119246	119345:	gap of unknown length
	* 119346	120636:	contig of 1291 bp in length
	* 120637	120736:	gap of unknown length
	* 120737	122861:	contig of 2125 bp in length
	* 122862	122961:	gap of unknown length
	* 122962	124308:	contig of 1347 bp in length
	* 124309	124408:	gap of unknown length
	* 124409	125524:	contig of 1116 bp in length
	* 125525	125624:	gap of unknown length
	* 125625	127500:	contig of 1876 bp in length
	* 127501	127600:	gap of unknown length
	* 127601	129895:	contig of 2295 bp in length
	* 129896	129995:	gap of unknown length
	* 129996	131463:	contig of 1468 bp in length
	* 131464	131563:	gap of unknown length
	* 131564	132883:	contig of 1320 bp in length
	* 132884	132983:	gap of unknown length
Query Match	Best Local Similarity	14.7%;	Score 88.2; DB 2; Length 162389;
Matches	99: Conservative	84.6%; Pred.No.	3.3e-10; Mismatches 18; Indels 0; Gaps 0;
OY	161 ggaatgcagagaagcctggaattcttcctcgacggctcgacagagatccaaagactgcaccaggag	220	
Dd	134892 gggttcccttggtgactgcgatcatcttcgaaaactgtgtaaaaaacgccagacctccacagag	134951	
OY	221 tttccggagcgcgcgtctgtagaagcccgcgcaagaagaagagatcctaataaacagtat	277	
Dd	134952 tcctgcagcgcctgcagaaagacgtccccaagacaagagaagcttatAACCACTTGGT	135008	
RESULT	% AF277591/C	3681 bp	mRNA linear ROD 24-JAN-2001
LOCUS	AF277591	spliced.	Mus musculus scurfin (Foxp3) mRNA, complete cds; alternatively spliced.
DEFINITION:	Af277591.1 GI:12407636		
ACCESSION	AF277591		
VERSION	AF277591.1		
KEYWORDS	house mouse,		
SOURCE	Mus musculus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 3681) Brunkow,M.E., Jeffrey,E.W., Hjertild,K.A., Paepfer,B., Clark,L.B., Vasylyk,S.A., Wilkinson,J.E., Galas,D., Ziegler,S.F. and Ramsdell,F. Disruption of a new forkhead/winged-helix protein, scurfin, results in the fatal lymphoproliferative disorder of the scurfy mouse Nat. Genet. 27 (1), 68-73 (2001)		
TITLE	JOURNAL MEDLINE PUBLISHED REFERENCE AUTHORS		
TITLE	2 (bases 1 to 3681) Brunkow,M.E., Jeffrey,E.W., Hjertild,K.A., Paepfer,B.W., Clark,L.B., Wilkinson,J.E., Ziegler,S.F. and Ramsdell,F. Direct Submission Submitted (15-JUN-2000) Genomics, Chromosome Rad Inc., 1631 220th		

APCPAVKSNACYWAVALLVFLNLTITASEHHGQFVWLTOQEVANKVLLCLFTVEML
LKLVCGPSAVSSEFNDFCVVCGGILETILVEVCAGVPLGIVLRCVRLRLPIKFK
TRHWSUNVASLNSKSTASLILLFLIFLIFSLGHLQEGGKFNFOITHKST
FDTFQALLIVFQILITGEDMWNWIDGMATGGPFPGLVLCVIFILFCGNYILLN
VFLAIVDNLASDAGTAKDKGKSEKMDQPPGNEGLVPGVEKEEGEERAGGADME
EEEEEEEEEDEEGAGGVELLOEVYPRKVVPIPEGSAFTCLSTNPLRKGCHTL
IHHVFTLILVILSVLSUAAEDPIRAHFRNHIILGYDAFTSITFVILLKHTV
FGALHRCGFCRSNFMNLDLAVSVSLISFGIHSSAISVVKILRLVRLPRAINRA
KGLHVQCVFVARTIGNIMVITLLOFNFAICGVOLFKEKFTCTDEAKHTPOECK
CSLVPDGVSRRLVRERLWNSDFHNVLSAMALFTVSTFEGWPAALLYKADAY
AEDGHPITNIRVEISFVYIIIAFTMMNIPVGVIIIFRAOGEQGEONCELDKNQ
POCEYALKAPQLRRYIPKPNHOFYVAVNSAAFYLMLLILNLVALANQHTFOT
APFNMYALNMVETGLTIEWLKIIIAFKPKHYETDAWNFTDALIVGSIQVIAVTE
VNMGHLESDEDSRISITIEFLRVHRLVKLSKGEIGIRILLMTIKFQFALPYVA
LLIAMIFILYIVIGMONGKVALDQTOINRNINFOITPQAVLLFRCACTGEANOEYL
LASLGRNCDPDSGPGFEETCSNFAIYFISFIMLCALILNLFVAVIMEDYL
TDSWILGPHLHDEKRIWSEYDGCAGRIKHLVDYALLRLOPLGFGKLCPIRVAC
KFLVAMNPLNSDGVITFNAILFALVITSLKIKTEGNDQANQELRIVIKKIKWEMKO

Query Match 20.2% Score 121.8; DB 9; Length 130788;
Best Local Similarity 67.1%; Pred. No. 5e-18;
Matches 233; Conservative 0; Mismatches 32; Indels 82; Gaps 1;
Oy 121 ggggtccatctctgaagtagtcagccactccgaagagctgcagattgcagagagctggaat 180
Db 55539 GGGGGCAATCCCTATGCTCTGCTACCTGCTCTGCTGTGGCTCATGCGAGCTGGAAT 55598
Oy 181 cttctcagcgcctggcagagatcaagaactgcaccagagatgtccggcgagctgctggaag 240
Db 55599 CTCTTCAGCGCTGGCAGAGATCAAGAACTGCACCAGAGTGTCCGGCGGCTGCTGGAAG 55659
Oy 241 agcccgagagagagagagctctataagcagct----- 274
Db 55659 AGGCTCGCAGGAGGAGAGGCTATTAAGCAGCTGTGAAGGCTGTGTGAGGACCTGGG 55718
Oy 275 -----gagtg 278
Db 55719 TAGCTTAGGAGGCTGGGGGATGGTCTCGGGCAGTGCCTGTATATCCCTGCTAGATG 55778
Oy 279 tcagagctgaagactctgccagagatgtgccggctgcctaccagcagcactcctg 338
Db 55779 TCAGAGCTGGAGACTCTGCCAGAGATGTGTCCGGCTGCGCTACACCCAGCGCATCCTG 55839
Oy 339 gagatcagcgaacatccagagacagagagagagagatcccaagat 385
Db 55839 GAGATGTGGGCAACATCCGAGACGACAGAGAGATACCAAGGT 55885

RESULT 7
AF277994/c 30858 bp DNA linear ROD 24-JAN-2001
LOCUS Mus musculus scurfin (Foxp3) gene, complete cds.
DEFINITION AF277994
ACCESSION AF277994
VERSION AF277994.1 GI:12407642
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
1 (bases 1 to 30858)
Brunkov,M.E., Jeffery,E.W., Hjerrild,K.A., Paepers,B., Clark,L.B.,
Yasayko,S.A., Wilkinson,J.E., Galas,D., Ziegler,S.F. and
Pansdell,F.
Disruption of a new forkhead/winged-helix protein, scurfin, results
in the fatal lymphoproliferative disorder of the scurfy mouse
Nat. Genet. 27 (1), 68-73 (2001)
20578751
11138001
2 (bases 1 to 30858)
Brunkov,M.E., Jeffery,E.W., Hjerrild,K.A., Paepers,B.W., Clark,L.B.,
Wilkinson,J.E., Ziegler,S.F. and Pansdell,F.
Direct Submission
Submitted (15-JUN-2000) Genomics, Chiroscience R&D Inc., 1631 220th

ST. SE. Bothell, WA 98021, USA
Location/Qualifiers
1. 30858
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/map="between DXMI55 and DXMIT161"
Join(12525..12678,18642..18724,19164..19391,19904..20008,
20105..20243,20617..20704,20803..20907,21598..21785,
21926..22006,24205..24358,24991..25067,25482..25583,
25732..28066)
/gene="Foxp3"
/note="alternatively spliced"
/product="scurfin"
12525..28066
/gene="Foxp3"
Join(12525..12678,18642..18724,19164..19184)
/gene="Foxp3"
/note="alternative"
12525..12678
/gene="Foxp3"
/number=-2b
13315..13410
/gene="Foxp3"
/number=-2a
Join(13315..13410,18642..18724,19164..19184)
/gene="Foxp3"
/note="alternative"
Join(13315..13410,18642..18724,19164..19391,19904..20008,
20105..20243,20617..20704,20803..20907,21598..21785,
21926..22006,24205..24358,24991..25067,25482..25583,
25732..28066)
/gene="Foxp3"
/note="alternatively spliced"
/product="scurfin"
18642..18724
/gene="Foxp3"
/number=-1
19164..19391
/gene="Foxp3"
/number=-1
Join(19185..19391,19904..20008,20105..20243,20617..20704,
20803..20907,21598..21785,21926..22006,24205..24358,
24991..25067,25482..25583,25732..25875)
/gene="Foxp3"
/codon_start=1
/product="scurfin"
/protein_id="AAG53608.1"
/db_xref="GI:12407642"
/translation="MPNRPAPKPMAPSLALGSPGVLPSPKATPKSGSELLTGRSGGP
FOGRDLRSGAHTSSSLNPLPSOLQPTVPLVAVPASCARLGSPHQLQLODSFHEM
HOLSTVDAAHOTPVLOVRPLDNPAIMSLPPSAATGVFSLKARGLPBGINVASLEWV
SREPALCTFPRSGTRKDSNLLAPOGSPPLLANGCKWPGCEKVEEPEELKHQ
ADHLIDKGAQCLLOREVQSLQOOLEKEKLGKMOAHLAGKNAKAPSVAMDR
SSCCIVATSTOGSVLPAMSPAREAPDGLFAVRHLNGSHGNSFPEFFHMDYFKYH
NMRFPTVAILRWAILLEAPERORTLNEIYHWFTRMFAYFRNHPATWKNATRRHLSLH
KCFVVESEKCAVWTVDEEFKRRSRQPNKCSNCP"
19904..20008
/gene="Foxp3"
/number=2
20105..20243
/gene="Foxp3"
/number=3
20617..20704
/gene="Foxp3"
/number=4
20803..20907
/gene="Foxp3"
/number=5
21698..21785
/gene="Foxp3"
/number=6
21926..22006


```

; TOPOLOGY: linear
US-09-225-170-13

Query Match 5.5%; Score 21; DB 3; Length 5162;
Best Local Similarity 100.0%; Pred.No.;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 cattaataataataataataa 382
|||||
DB 4859 CATTAAAAAATATAAAAAAA 4839

RESULT 26
US-08-441-139-12/c
; Sequence 12, Application US/08441139
; Patent No. 5773245
; GENERAL INFORMATION:
; APPLICANT: Wittrup, Dr. Karl D.
; APPLICANT: Robinson, Anne S.
; TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF
; TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441.139
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/089,997
; FILING DATE: 06-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8646
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5470 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: exon
; LOCATION: 593..715
; FEATURE:
; NAME/KEY: exon
; LOCATION: 806..1036
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1402..1539
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2175..2289
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2378..2764
; FEATURE:

```



```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/070,301
;   FILING DATE: 24-MAY-1991
;   CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP 1-209687
;   FILING DATE: 15-AUG-1989
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP 1-181933
;   FILING DATE: 31-OCT-1985
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP 2-76331
;   FILING DATE: 26-MAR-1990
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP 2-106412
;   FILING DATE: 24-APR-1990
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP 2-205475
;   FILING DATE: 02-AUG-1990
; ATTORNEY/AGENT INFORMATION:
;   NAME: Player, William E.
;   REGISTRATION NUMBER: 31,409
;   REFERENCE/DOCKET NUMBER: P-450-22830
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 887-040
;   TELEFAX: (202) 835-0605
;   TELEX: 440706
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 6638 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
;   ORGANISM: Horse
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 11..3070
; US-08-070-301-2

      Query Match          5.5%; Score 21; DB 2; Length 6638;
      Best Local Similarity 100.0%; Pred. No. 1.9;
      Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 cattaataataataataataataa 382
      |||
Db 5186 CATTAAAAAATAAAAAAAAAA 5206

RESULT 29
US-09-318-448-11/c
; Sequence 11, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
;   APPLICANT: Johnson, William G.
;   APPLICANT: Stenroos, Edward S.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318,448
; CURRENT FILING DATE: 1999-03-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 18596
; TYPE: DNA

; ORGANISM: Homo sapiens
US-09-318-448-11

      Query Match          5.5%; Score 21; DB 1; Length 20303;
      Best Local Similarity 100.0%; Pred. No. 1.8;
      Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 cattaataataataataataataa 382
      |||
Db 13143 CATTAAAAAATAAAAAAAAAA 13123

RESULT 31
US-08-370-975B-1/c
; Sequence 1, Application US/08370975B
; Patent No. 5622851
; GENERAL INFORMATION:
;   APPLICANT: Maley, Frank
;   APPLICANT: Gladys F.
;   APPLICANT: Weiner, Karen X.B.
; TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Nixon, Hargrave, Devans & Doyle
;   STREET: Clinton Square, P.O. Box 1051
;   CITY: Rochester
;   STATE: New York
;   COUNTRY: USA
;   ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/370,975B
;   FILING DATE: 10-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;   NAME: Timian, Susan J.
;   REGISTRATION NUMBER: 34,103
;   REFERENCE/DOCKET NUMBER: 20894/80
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (716)263-1636
;   TELEFAX: (716)263-1600
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 20303 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; POSITION IN GENOME:
;   CHROMOSOME/SEGMENT: 4q35
; US-08-370-975B-6

      Query Match          5.5%; Score 21; DB 1; Length 20303;
      Best Local Similarity 100.0%; Pred. No. 1.8;
      Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 cattaataataataataataataa 382
      |||
Db 13143 CATTAAAAAATAAAAAAAAAA 13123

RESULT 31
US-08-370-975B-1/c
; Sequence 1, Application US/08370975B
; Patent No. 5622851
; GENERAL INFORMATION:
;   APPLICANT: Maley, Frank
```

APPLICANT: Maley, Gladys F.
APPLICANT: Weiner, Karen X.B.
TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: USA
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/370,975B
FILING DATE: 10-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Timian, Susan J.
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 20894/80
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716)263-1636
TELEFAX: (716)263-1600
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 26764 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 4q35
US-08-370-975B-1

Query Match 5.5%; Score 21; DB 1; Length 26764;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 362 cattataaaaaaaaaaaaaa 382
Db 15106 cattataaaaaaaaaaaaaa 15086

RESULT 32
US-08-113-646A-6
Sequence 6, Application US/08113646A
Patent No. 5578468
GENERAL INFORMATION:
APPLICANT: PICKUP, David J.
APPLICANT: PATEL, Dhaval Kumar
APPLICANT: ANTICZAK, James B.
TITLE OF INVENTION: SITE-SPECIFIC RNA CLEAVAGE
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/113,646A
FILING DATE: 31-AUG-1993

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/084,406
FILING DATE: 10-AUG-1987
ATTORNEY/AGENT INFORMATION:
NAME: WILSON, MARY J.
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 1579-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 815-4000
TELEFAX: (703) 816-4100
TELEX: 200757 MIXX UR
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 36 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA (genomic)
US-08-113-646A-6

Query Match 5.2%; Score 20; DB 1; Length 36;
Best Local Similarity 90.0%; Pred. No. 7.7;
Matches 18; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 363 attataaaaaaaaaaaaaa 382
Db 16 AUUAAAAAAAAAAAAAAAAA 35

RESULT 33
US-08-799-464A-35
Sequence 35, Application US/08799464A
Patent No. 5958601
GENERAL INFORMATION:
APPLICANT: Murtaugh, Michael P. et al.
TITLE OF INVENTION: VR-2332 VIRAL NUCLEOTIDE SEQUENCE AND METHODS OF USE
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: John M. Collins
STREET: 2405 Grand Blvd., Suite 400
CITY: Kansas City
STATE: Missouri
COUNTRY: USA
ZIP: 64108
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,464A
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/287,941
FILING DATE: August 5, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26,262
REFERENCE/DOCKET NUMBER: 22907
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816) 474-9050
TELEFAX: (816) 474-9057
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 135 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA

```

: HYPOTHETICAL: NO
: US-08-799-464A-35

Query Match          5.2%; Score 20; DB 2; Length 135;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 attaaaaa382
: 11111111111111111111
Db 113 ATTAAAAA382

RESULT 34
US-08-799-464A-34
: Sequence 34, Application US/08799464A
: Patent No. 5998601
: GENERAL INFORMATION:
: APPLICANT: Mutaugh, Michael P. et al.
: TITLE OF INVENTION: VR-2332 VIRAL NUCLEOTIDE SEQUENCE AND
: METHODS OF USE
: NUMBER OF SEQUENCES: 35
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: John M. Collins
: STREET: 2405 Grand Blvd., Suite 400
: CITY: Kansas City
: STATE: Missouri
: COUNTRY: USA
: ZIP: 64108
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/799,464A
: FILING DATE:
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/287,941
: FILING DATE: August 5, 1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Collins, John M.
: REGISTRATION NUMBER: 26,262
: REFERENCE/DOCKET NUMBER: 22907
: TELEPHONE: (816) 474-9050
: TELEFAX: (816) 474-9057
: INFORMATION FOR SEQ ID NO: 34:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 171 bases
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
: HYPOTHETICAL: NO
: US-08-799-464A-34

Query Match          5.2%; Score 20; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 attaaaaa382
: 11111111111111111111
Db 149 ATTAAAAA168

RESULT 35
US-08-721-488-9
: Sequence 9, Application US/08721488
: Patent No. 5965388
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John
: APPLICANT: LaVallie, Edward
: APPLICANT: Racie, Lisa
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Spaulding, Vikki
: APPLICANT: Bowman, Michael
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
: TITLE OF INVENTION: ENCODING THEM
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 CambridgePark Drive
: CITY: Cambridge
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/721,488
: FILING DATE:
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Brown, Scott A.
: REGISTRATION NUMBER: 32,724
: TELEPHONE: (617) 498-8224
: TELEFAX: (617) 876-5651
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 358 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-721-488-9

Query Match          5.2%; Score 20; DB 2; Length 358;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 attaaaaa382
: 11111111111111111111
Db 337 ATTAAAAA356

RESULT 36
US-08-117-080-13
: Sequence 13, Application US/08117080
: Patent No. 5482928
: GENERAL INFORMATION:
: APPLICANT: DE BOLLE, MIGUEL
: APPLICANT: BROEKERT, WILLEM F
: APPLICANT: CAMMUE, BRUNO PA
: APPLICANT: VANDERLEYDEN, JOZEF
: APPLICANT: REES, SARAH B
: TITLE OF INVENTION: BIOCIDAL PROTEINS
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CUSHMAN DARBY & CUSHMAN
: STREET: 1100 NEW YORK AVENUE, NW; NINTH FLOOR, EAST
: CITY: WASHINGTON
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3918
: COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/117,080
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB92/00423
; FILING DATE: 10-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9105052.6
; FILING DATE: 11-MAR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9105684.6
; FILING DATE: 19-MAR-1991
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861 3000
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: FIGURE 7 BASE SEQUENCE MJ-AMP2
US-08-117-080-13

Query Match 5.2%; Score 20; DB 1; Length 433;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 attaaaaaiaaaaaaiaaaaa 382
DB 413 ATTAAAAAIAAAAAAAAAA 432

RESULT 37
US-08-471-329-13
; Sequence 13, Application US/08471329
; Patent No. 5685048
; GENERAL INFORMATION:
; APPLICANT: DE BOLLE, MIGUEL
; APPLICANT: BROEKAERT, WILLEM F
; APPLICANT: CAMMUE, BRUNO PA
; APPLICANT: VANDERLEYDEN, JOZEF
; APPLICANT: REES, SARAH B
; TITLE OF INVENTION: BIOCIDAL PROTEINS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DABY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, NW; NINTH FLOOR, EAST
; STREET: TOWER
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,329
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/117,080

; FILING DATE: 20-DEC-1993
; APPLICATION NUMBER: PCT/GB92/00423
; FILING DATE: 10-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9105052.6
; FILING DATE: 11-MAR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9105684.6
; FILING DATE: 19-MAR-1991
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861 3000
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: FIGURE 7 BASE SEQUENCE MJ-AMP2
US-08-471-329-13

Query Match 5.2%; Score 20; DB 1; Length 433;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 attaaaaaiaaaaaaiaaaaa 382
DB 413 ATTAAAAAIAAAAAAAAAA 432

RESULT 38
US-08-915-142-13
; Sequence 13, Application US/08915142
; Patent No. 5942663
; GENERAL INFORMATION:
; APPLICANT: DE BOLLE, MIGUEL
; APPLICANT: BROEKAERT, WILLEM F
; APPLICANT: CAMMUE, BRUNO PA
; APPLICANT: VANDERLEYDEN, JOZEF
; APPLICANT: REES, SARAH B
; TITLE OF INVENTION: BIOCIDAL PROTEINS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DABY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, NW; NINTH FLOOR, EAST
; STREET: TOWER
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,142
; FILING DATE: 20-AUG-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/471,329
; FILING DATE: 02-JUN-1995
; APPLICATION NUMBER: US 08/117,080
; FILING DATE: 20-DEC-1993
; APPLICATION NUMBER: PCT/GB92/00423
; FILING DATE: 10-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9105052.6
; FILING DATE: 11-MAR-1991

PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9105684.6
FILING DATE: 19-MAR-1991
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861 3000
TELEFAX: (202) 822-0944
TELEX: 6714677 CUSH
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 433 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: FIGURE 7 BASE SEQUENCE MJ-AMP2
US-08-915-142-13

Query Match 5.2%; Score 20; DB 2; Length 433;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 attaaaaa382
DB 413 ATTAATAAAAAAAAAAAAAA 432

RESULT 39

US-08-289-247B-2
Sequence 2, Application US/08289247B
Patent No. 5728579

GENERAL INFORMATION:
APPLICANT: Morrison, Briggs W.
APPLICANT: Leder, Philip
TITLE OF INVENTION: Detection and Treatment of Breast
TITLE OF INVENTION: Cancer
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2223

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/289,247B

FILING DATE: August 11, 1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Bleker-Brady, Kristina

REGISTRATION NUMBER: 39,109

REFERENCE/DOCKET NUMBER: 00383/021001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 428-0200

TELEFAX: (617) 428-7045

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 526

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-289-247B-2

Query Match 5.2%; Score 20; DB 1; Length 526;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 attaaaaa382
DB 504 ATTAATAAAAAAAAAAAAAA 523

RESULT 40

PCT-US95-09762-2

Sequence 2, Application PC/TUS9509762

GENERAL INFORMATION:

APPLICANT: Morrison, Briggs W.

APPLICANT: Leder, Philip

TITLE OF INVENTION: Detection and Treatment of Breast

TITLE OF INVENTION: Cancer

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson

STREET: 225 Franklin Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

COMPUTER: IBM PS/2 Model 502 or 55SX

OPERATING SYSTEM: MS-DOS (Version 5.0)

SOFTWARE: WordPerfect (Version 5.1)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/09762

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/289,247

FILING DATE: August 11, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Clark, Paul T.

REGISTRATION NUMBER: 30,162

REFERENCE/DOCKET NUMBER: 00383/021001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 526

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

PCT-US95-09762-2

Query Match 5.2%; Score 20; DB 5; Length 526;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 attaaaaa382
DB 504 ATTAATAAAAAAAAAAAAAA 523

RESULT 41

US-09-040-984-70

Sequence 70, Application US/09040984

Patent No. 6210883

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Wang, Tonglong

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS

TITLE OF INVENTION: OF LUNG CANCER

NUMBER OF SEQUENCES: 86

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040,984
FILING DATE: 18-MAR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.456
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-282-6031
TELEX:
INFORMATION FOR SEQ ID NO: 70:
SEQUENCE CHARACTERISTICS:
LENGTH: 536 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-040-984-70

Query Match 5.2% Score 20; DB 4; Length 536;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 attaaaaaiaaaaaaaaaa 382
|||||iiiiiiiiiiiiii
DB 513 ATTAATAAAAAAAAAAAAAA 532

RESULT 42
US-09-328-111-68
Sequence 68, Application US/09328111
Patent No. 6262333
GENERAL INFORMATION:
APPLICANT: Endege, Wilson O.
APPLICANT: Steinmann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll III, Eddie
APPLICANT: Dertli, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328.111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 68
LENGTH: 579
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:

NAME/KEY: misc feature
LOCATION: (1)_(579)
OTHER INFORMATION: n ~ A,T,C or G
US-09-328-111-68

Query Match 5.2% Score 20; DB 4; Length 579;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 attaaaaaiaaaaaaaaaa 382
|||||iiiiiiiiiiiiii
DB 191 attaaaaaiaaaaaaaaaa 210

RESULT 43
US-08-229-393-1
Sequence 1, Application US/08229393
Patent No. 5591434
GENERAL INFORMATION:
APPLICANT: Jenkins, Mark C
APPLICANT: Fayer, Ronald
APPLICANT: Tilley, Michael
APPLICANT: Upton, Steven J
TITLE OF INVENTION: DNA Sequence Encoding Surface Protein of
Patent No. 5591434
TITLE OF INVENTION: Cryptosporidium Parvum
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: USDA-ARS-OFT-NCAUR
STREET: 1815 N. University Street
CITY: Peoria
STATE: IL
COUNTRY: U.S.A.
ZIP: 61604
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/229,393
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Deck, Randal E.
REGISTRATION NUMBER: 34,078
REFERENCE/DOCKET NUMBER: P.C. 0108.94
TELECOMMUNICATION INFORMATION:
TELEPHONE: (309)-681-6515
TELEFAX: (309)-681-6688
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 602 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Cryptosporidium parvum
DEVELOPMENTAL STAGE: Sporozoite
IMMEDIATE SOURCE:
CLONE: CP15/60
FEATURE:
NAME/KEY: CDS
LOCATION: 1..474
US-08-229-393-1

Query Match 5.2% Score 20; DB 1; Length 602;
Best Local Similarity 100.0%; Pred. No. 6.1;

; ADDRESS: SHERIDAN ROSS F.C.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2002, 10:50:11 ; Search time 728.83 Seconds
(without alignments)
708.135 Million cell updates/sec

Title: US-09-248-178-60
Perfect score: 602
Sequence: 1 tgaagagccgcggtgag.....tcatgcggaggttcgagac 602

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 10

Total number of hits satisfying chosen parameters: 284374

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N.Geneseq_ll01.*
1: /SIDSI/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq/NA2001.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	602	100.0	602	20 AAX84206	DNA encoding human
2	602	100.0	602	21 AAC79435	CDNA sequence of h
3	456	75.7	1815	21 AAC78023	Human cancer assoc
4	18	3.0	459	22 AAF83407	P. chrysogenum ABC
5	18	3.0	459	22 AAF83408	Plant viral moveme
6	18	3.0	771	21 AAC79351	Human leukocyte ce
7	18	3.0	1113	19 AAV49553	Human prostate can
8	18	3.0	1557	20 AAZ33458	C. thermoaminogene
9	18	3.0	1704	22 AAH45373	DNA encoding human
10	18	3.0	1770	19 AAV38502	Human EST-derived
11	18	3.0	1776	22 AAH98783	

12	18	3.0	1824	19 AAV49552	Human leukocyte ce
13	18	3.0	2655	22 AAD05226	Human secreted pro
14	18	3.0	2271	15 AAQ62176	Human TLE-2 gene.
15	18	3.0	2271	15 AAQ45334	Human TLE-2 gene.
16	18	3.0	7746	21 AAA07836	Human homologue of
17	17	2.8	458	20 AAX40583	Human secreted pro
18	17	2.8	385	21 AAC06962	Human secreted pro
19	17	2.8	413	21 AAC06983	Human ORFX ORF1727
20	17	2.8	477	21 AAC76172	Leek glossy8 (gl8)
21	17	2.8	872	21 AAA57874	Leek glossy8 (gl8)
22	17	2.8	872	21 AAA57875	Phosphatidylinosit
23	17	2.8	951	22 AAH50961	Human G protein-co
24	17	2.8	1020	21 AAA30601	DNA encoding human
25	17	2.8	1020	21 AAA30718	S. spinosa DNA fra
26	17	2.8	1194	22 AAF88323	Human secreted pro
27	17	2.8	1258	21 AAC93450	Seven transmembran
28	17	2.8	1901	15 AAC66178	Human R12 seven tr
29	17	2.8	1901	19 AAV18356	Human 7TM receptor
30	17	2.8	1901	21 AAAG1725	Sequence encoding
31	17	2.8	1985	7 AAN60126	Human CDNA sequenc
32	17	2.8	2010	22 AAH14321	Human G-protein th
33	17	2.8	2453	18 AAT44092	C glutamicum codin
34	17	2.8	2766	22 AAH66941	Human ORFX ORF1165
35	17	2.8	2813	21 AAC75611	Human polynucleoti
36	17	2.8	3092	22 AA160348	Human bone marrow
37	17	2.8	3092	22 AAH90050	Human polynucleoti
38	17	2.8	3352	22 AA158562	Human bone marrow
39	17	2.8	3405	22 AAH89937	Human ORFX ORF1915
40	17	2.8	5713	21 AAC76360	Human liver cell s
41	17	2.8	5813	22 AAH57490	Human hepatocyte n
42	17	2.8	6411	19 AAV52730	Streptococcus pneu
43	17	2.8	16535	19 AAV52207	S. spinosa DNA fra
44	17	2.8	45824	22 AAF88315	S. spinosa DNA fra
45	17	2.8	50000	22 AAF88312	

ALIGNMENTS

RESULT 1
AAX84206
ID AAX84206 standard; cDNA; 602 BP.
XX
XX AAX84206;
AC
XX 08-SEP-1999 (first entry)
DI
XX
DE DNA encoding human breast tumour protein immunogenic fragment.
XX
XX Breast tumour protein; immunogenic fragment; vaccine; detection;
KW Breast cancer development; therapy; ss.
XX
XX Homo sapiens.
OS
XX
XX W05933869-A2.
XX
XX 08-JUL-1999.
XX
XX 22-DEC-1998; 98WO-US/7416.
XX
XX 17-JUL-1998; 98US-018627.
PR 24-DEC-1997; 97US-0998253.
PR 24-DEC-1997; 97US-0998255.
PR 17-JUL-1998; 98US-0118554.
XX
XX (CORI-) CORIYA CORP.
XX
XX Reed SG, Xu J;
XX
XX WPI; 1999-405486/34.
XX
XX New breast tumour protein genes used, in vaccines for immunotherapy,
PR or for diagnosis of breast cancer

XX PS Claim 3; Page 59; 70pp; English.

CC This sequence encodes a human breast tumour protein immunogenic fragment

CC of the invention. The polypeptides or nucleic acids encoding them are

CC useful in vaccines and pharmaceutical compositions for manufacture of

CC medicaments for inhibiting the development of breast cancer in a patient.

CC They can also be used to treat breast cancer. Antibodies against these

CC polypeptides can be used to detect and monitor progression of breast

CC cancer in patients. Primers and probes derived from the polynucleotides

CC encoding the breast proteins are useful for detection of breast cancer.

CC Peripheral blood cells from a patient incubated in the presence of at

CC least one polypeptide, such that T cells proliferate, are useful in

CC manufacture of a medicament for treating breast cancer in a patient.

CC Antigen presenting cells incubated in the presence of at least one

CC polypeptide are also useful for treating breast cancer.

XX SQ Sequence 602 BP; 147 A; 155 C; 192 G; 108 T; 0 Other;

Query Match 100.0%; Score 602; DB 20; Length 602;

Best Local Similarity 100.0%; Pred. No. 9,4e-289;

Matches 602; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgaagagccgcggtgagctgctgcccgatggagactgccaaccttcccaagctgcagc 60

DB 1 tgaagagccgcggtgagctgctgcccgatggagactgccaaccttcccaagctgcagc 60

QY 61 ttgtgtgagaatagtcgccagcgggttcaccttggcgggtcagtgaggagaagcacc 120

DB 61 ttgtgtgagaatagtcgccagcgggttcaccttggcgggtcagtgaggagaagcacc 120

QY 121 ggtccatctctgtgagctacccagcctccaaagctcagattgcagagctggagt 180

DB 121 ggtccatctctgtgagctacccagcctccaaagctcagattgcagagctggagt 180

QY 181 ctctcagcggctggcagagctgccaagactgccaagctgccaagctgccaag 240

DB 181 ctctcagcggctggcagagctgccaagactgccaagctgccaagctgccaag 240

QY 241 actccctatctgggaagctggacggacgttttcgggtgactgagctgtgttcaagg 480

DB 241 actccctatctgggaagctggacggacgttttcgggtgactgagctgtgttcaagg 480

QY 481 atgccagaagagacgtgctgttcggaagccctataagctatcagctgtctcagcaga 540

DB 481 atgccagaagagacgtgctgttcggaagccctataagctatcagctgtctcagcaga 540

QY 541 actcagcagctctatccagaccatcagagacacagcagccaccatcctcggtggttcgag 600

DB 541 actcagcagctctatccagaccatcagagacacagcagccaccatcctcggtggttcgag 600

QY 601 ac 602

DB 601 ac 602

AC AAC75435;

XX 07-FEB-2001 (first entry)

LI cDNA sequence of human breast tumour clone 1016L6.

DE Human; breast tumour antigen; cytostatic; immunotherapy;

XX breast cancer; vaccine; ss.

OS Homo sapiens.

XX WO200061756-A2.

XX 19-OCT-2000.

XX 10-APR-2000; 2000WO-US0668.

XX 09-APR-1999; 59US-020950.

XX 02-JUL-1999; 59US-0346327.

XX (CORI-) CORIXA CORP.

XX Reed SG, Xu J, Dillman DC;

XX WPI: 2000-636556/61.

XX A novel isolated polypeptide comprising an immunogenic portion of a breast cancer protein useful in the detection and treatment of breast cancer.

XX Claim 4; Page 75-76; y5pp; English.

XX The present sequence was isolated from a breast tumour cDNA library. It is provided in a specification relating to compounds for immunotherapy and diagnosis of breast cancer. Breast tumour antigens and the polynucleotides that encode them may be used in the production of a pharmaceutical composition to be used in the treatment of breast cancer. Proliferated T cells and incubated antigen presenting cells are also required. The polypeptides and polynucleotides may also be used to produce a vaccine.

XX SQ Sequence 602 BP; 147 A; 155 C; 192 G; 108 T; 0 Other;

Query Match 100.0%; Score 602; DB 21; Length 602;

Best Local Similarity 100.0%; Pred. No. 9,4e-289;

Matches 602; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgaagagccgcggtgagctgctgcccgatggagactgccaaccttgcgaagctgcagc 60

DB 1 tgaagagccgcggtgagctgctgcccgatggagactgccaaccttgcgaagctgcagc 60

QY 61 ttgtgtgagaatagtcgccagcgggttcaccttggcgggtcagtgaggagaagcacc 120

DB 61 ttgtgtgagaatagtcgccagcgggttcaccttggcgggtcagtgaggagaagcacc 120

QY 121 ggtccatctctgtgagctacccagcctccaaagctcagattgcagagctggagt 180

DB 121 ggtccatctctgtgagctacccagcctccaaagctcagattgcagagctggagt 180

QY 181 ctctcagcggctggcagagctgccaagactgccaagctgccaagctgccaag 240

DB 181 ctctcagcggctggcagagctgccaagactgccaagctgccaagctgccaag 240

QY 241 actccctatctgggaagctggacggacgttttcgggtgactgagctgtgttcaagg 480

DB 241 actccctatctgggaagctggacggacgttttcgggtgactgagctgtgttcaagg 480

QY 481 atgccagaagagacgtgctgttcggaagccctataagctatcagctgtctcagcaga 540

DB 481 atgccagaagagacgtgctgttcggaagccctataagctatcagctgtctcagcaga 540

QY 541 actcagcagctctatccagaccatcagagacacagcagccaccatcctcggtggttcgag 600

DB 541 actcagcagctctatccagaccatcagagacacagcagccaccatcctcggtggttcgag 600

QY 601 ac 602

DB 601 ac 602

Db 361 agcagaagagagatcaccaagattgtctgatacgaaggagctcagaagaaatca 420
 Qy 421 actcctctctggaagctgacccgacgtttgctgctgactgagctgtgttcaagg 480
 Db 421 actcctctctggaagctgacccgacgtttgctgctgactgagctgtgttcaagg 480
 Qy 481 atgcaagaagagagatgcttctcgaagcctataagctatctagctgctgacaga 540
 Db 481 atgcaagaagagagatgcttctcgaagcctataagctatctagctgctgacaga 540
 Qy 541 actgcagcagctcatccagaccatcagagacacagcagccaccatctcggaggttcgag 600
 Db 541 actgcagcagctcatccagaccatcagagacacagcagccaccatctcggaggttcgag 600
 Qy 601 ac 602
 Db 601 ac 602

RESULT 3

AACT8023
 ID AACT8023 standard; cDNA; 1815 BP.

XX AACT8023;

XX 08-FEB-2001 (first entry)

XX Human cancer associated gene sequence SEQ ID NO:417.

XX Human; cancer associated gene; cancer antigen; detection; cancer;
 diagnosis; cytostatic; proliferative; vulnary; immunomodulator;
 antiinflammatory; antitumor; antirheumatic; antithrombotic; antiviral;
 antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
 dermatological; neuroprotective; thrombolytic; coagulant; neotropic;
 vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
 immune disorder; haematopoietic cell disorder; autoimmune disorder;
 allergic reaction; graft versus host disease; organ rejection;
 haemostatic; thrombolytic; cardiovascular disorder; infection;
 neurological disease; drug screening; ss.

XX Homo sapiens.

XX WO200055350-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05882.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2000-587533/55.

XX P-PSDB; AAB43814.

XX Novel isolated nucleic acids comprising sequences encoding peptides
 useful for treating or diagnosing e.g. cancer

XX Claim 1; Page 958-959; 2352pp; English.

XX AACT7607 to AACT78448 encode the human cancer associated proteins given
 in AAB43398 to AAB44239. The proteins can have activities based on the
 tissues and cells the genes are expressed in. Example of activities
 include: cytostatic; proliferative; vulnary; immunomodulator;
 antiinflammatory; antitumor; antirheumatic; antithrombotic;
 antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
 dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
 neotropic; vasotropic; antipsoriatic and antiangiogenic. The
 polynucleotides and polypeptides can be used for preventing, treating or
 ameliorating medical conditions and diagnosing pathological conditions.

CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
 CC the present invention may be used to treat immune disorders by activating
 CC or inhibiting the proliferation, differentiation or mobilisation of
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune
 CC disorders, allergic reactions, graft versus host disease and organ
 CC rejection, modulate haemostatic or thrombolytic activity, modulate
 CC inflammation, cancers, cardiovascular disorders, neurological disease an:
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,
 CC agonists and antagonists may be also be used in drug screens. AACT78448 to
 CC AACT78457 and AAB44240 represent sequences used in the exemplification of
 CC the present invention.

XX Sequence 1815 BP; 372 A; 477 C; 613 G; 350 T; 3 other;

Query Match 75.7%; Score 456; DB 21; Length 1815;

Best Local Similarity 100.0%; Pred. No. 2.2e-216;

Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 147 ctccgaagctgcagagattcagagagctggaattctctcgcagcgtcggcagatccaa 206
 Db 599 ctccgaagctgcagagattcagagagctggaattctctcgcagcgtcggcagatccaa 658

Qy 207 gaactgcaccagaggtctccggcggtcgtggaagagcccgaggaaggaggtctat 266
 Db 655 gaactgcaccagaggtctccggcggtcgtggaagagcccgaggaaggaggtctat 718

Qy 267 aagcagctgagtcagagctgaggagactctgccagagatgtgtcccggtcggctacacc 326
 Db 719 aagcagctgagtcagagctgaggagactctgccagagatgtgtcccggtcggctacacc 778

Qy 327 cagcgcctctgcagctcgtggcacaatccggaagcagaggaagagatccacaaatc 386
 Db 779 cagcgcctctgcagctcgtggcacaatccggaagcagaggaagagatccacaaatc 838

Qy 387 ttgtctgatacgaagagcttcagaagaatacctccctctctctggaagctggaccgg 446
 Db 839 ttgtctgatacgaagagcttcagaagaatacctccctctctctggaagctggaccgg 898

Qy 447 acgtttgcgtgactgactgactgtgttcaagagatcccaagagacgagctgtctcg 506
 Db 899 acgtttgcgtgactgactgactgtgttcaagagatcccaagagacgagctgtctcg 958

Qy 507 aagcctctataagttatctgactgctctgcagagaaactgcagcagctcaccagaccatc 566
 Db 959 aagcctctataagttatctgactgctctgcagagaaactgcagcagctcaccagaccatc 1018

Qy 567 gaggacacagggcaccacatcgcgggaggttcagagac 602
 Db 1019 gaggacacagggcaccacatcgcgggaggttcagagac 1054

RESULT 4

AAFB3407/C

ID AAFB3407 standard; cDNA; 459 BP.

XX AAFB3407;

XX 09-JUL-2001 (first entry)

XX P. chrysogenum ABC transporter dd062 partial cDNA.

XX Beta-lactam; micro-organism; ABC transporter; ATP-binding cassette;
 KW adenosine triphosphate-binding cassette; antibiotic; penicillin;
 KW cephalosporin; dd062; ss.

XX Penicillium chrysogenum.

XX WO200132904-A2.

XX 10-MAY-2001.

XX 03-NOV-2000; 2000WO-EP11469.

XX 03-NOV-1999; 99EP-0203684.
 PR 03-NOV-1999; 99EP-0203685.
 PR 03-NOV-1999; 99EP-0203687.
 PR 03-NOV-1999; 99EP-0203688.
 PR 03-NOV-1999; 99EP-0203689.
 PR 03-NOV-1999; 99EP-0203690.
 PR 03-NOV-1999; 99EP-0203691.
 PR 03-NOV-1999; 99EP-0203692.
 PR 03-NOV-1999; 99EP-0203693.
 PR 03-NOV-1999; 99EP-0203694.
 XX (STAM) DSM NV.
 PA
 XX Van Den Berg MA, Bovenberg RAL, Driessen AJM, Konings WN;
 PI Schuurs TA, Nieboer M, Westerlaken I;
 XX WPI; 2001-291055/30.
 DR
 DR P-PSDB; AAB62495.
 XX
 XX Enhancing secretion of beta-lactam compounds from a micro-organism by
 PT enhancing adenosine triphosphate-binding cassette transporter activity,
 PT useful for producing e.g. penicillin and cephalosporins -
 XX
 PS Claim 9; Page 116; 116pp; English.
 XX
 CC The invention relates to a method for enhancing the secretion of beta-
 CC lactam compounds from a micro-organism that comprises enhancing ABC
 CC (adenosine triphosphate (ATP)-binding cassette) transporter activity of
 CC the micro-organism. The method is used for enhancing the production and
 CC secretions of beta-lactam antibiotics like penicillin and cephalosporins.
 CC Manipulation of the ABC-transport protein system provides a means for
 CC enhancing beta-lactam secretion. The present sequence represents a
 CC partial cDNA sequence of the P. chrysogenum ABC transporter dd062.
 XX
 SQ Sequence 459 BP; 100 A; 133 C; 119 G; 107 T; 0 other;

Query Match 3.0%; Score 18; DB 22; Length 459;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 534 cactgagaactgcagccag 551
 |||||
 Db 160 CACGAGAACTGCAGCCAG 143

RESULT 5
 AAF83408/c
 ID AAF83408 standard; cDNA; 459 BP.
 XX
 AC AAF83408;
 XX
 DT 09-JUL-2001 (first entry)
 XX
 DE P. chrysogenum ABC transporter dd062 cDNA.
 XX
 KW Beta-lactam; micro-organism; ABC transporter; ATP-binding cassette;
 KW adenosine triphosphate-binding cassette; antibiotic; penicillin;
 KW cephalosporin; dd062; ss.
 XX
 OS Penicillium chrysogenum.
 XX
 XX Key Location/Qualifiers
 FH CDS 4..457
 FT /*tag= a
 FT /product= "dd062"
 FT /note= "the start and stop codons are not indicated"
 XX
 PN WO200132904-A2.
 XX
 PD 10-MAY-2001.
 XX
 PF 03-NOV-2000; 2000WO-EP11489.

XX 03-NOV-1999; 99EP-0203684.
 PR 03-NOV-1999; 99EP-0203685.
 PR 03-NOV-1999; 99EP-0203687.
 PR 03-NOV-1999; 99EP-0203688.
 PR 03-NOV-1999; 99EP-0203689.
 PR 03-NOV-1999; 99EP-0203690.
 PR 03-NOV-1999; 99EP-0203691.
 PR 03-NOV-1999; 99EP-0203692.
 PR 03-NOV-1999; 99EP-0203693.
 PR 03-NOV-1999; 99EP-0203694.
 XX (STAM) DSM NV.
 PA
 XX Van Den Berg MA, Bovenberg RAL, Driessen AJM, Konings WN;
 PI Schuurs TA, Nieboer M, Westerlaken I;
 XX WPI; 2001-291055/30.
 DR
 DR P-PSDB; AAB62495.
 XX
 XX Enhancing secretion of beta-lactam compounds from a micro-organism by
 PT enhancing adenosine triphosphate-binding cassette transporter activity,
 PT useful for producing e.g. penicillin and cephalosporins -
 XX
 PS Claim 9; Page 116-117; 116pp; English.
 XX
 CC The invention relates to a method for enhancing the secretion of beta-
 CC lactam compounds from a micro-organism that comprises enhancing ABC
 CC (adenosine triphosphate (ATP)-binding cassette) transporter activity of
 CC the micro-organism. The method is used for enhancing the production and
 CC secretions of beta-lactam antibiotics like penicillin and cephalosporins.
 CC Manipulation of the ABC-transport protein system provides a means for
 CC enhancing beta-lactam secretion. The present sequence represents the
 CC cDNA sequence of the P. chrysogenum ABC transporter dd062.
 XX
 SQ Sequence 459 BP; 100 A; 133 C; 119 G; 107 T; 0 other;

Query Match 3.0%; Score 18; DB 22; Length 459;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 534 cactgagaactgcagccag 551
 |||||
 Db 160 CACGAGAACTGCAGCCAG 143

RESULT 6
 AAC79351
 ID AAC79351 standard; cDNA; 771 BP.
 XX
 AC AAC79351;
 XX
 DT 06-FEB-2001 (first entry)
 XX
 DE Plant viral movement protein encoding cDNA SEQ ID 7.
 XX
 KW Plant viral movement protein; transport; transgenic plant;
 KW viral resistance; cosuppression; ss.
 XX
 OS Hevea brasiliensis.
 XX
 PN WO2000060088-A2.
 PD 12-OCT-2000.
 XX
 PF 06-APR-2000; 2000WO-US05110.
 XX
 PR 07-APR-1999; 59US-0128092.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Krebbers E, Weng 2, Cahoon RE;

Tue Apr 30 14:18:49 2002

us-09-248-178-60.rng

CC specificity). This approach eliminates many of the false results, as
 CC regards tissue specificity, associated with known methods that use
 CC single (usually short) ESTs. AA233423-233476 represent expressed
 CC sequence tags described in the method of the invention.

XX
 SQ Sequence 1557 BP; 293 A; 538 C; 398 G; 328 T; 0 other;

Query Match
 Best Local Similarity 3.0%; Score 18; DB 20; Length 1557;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 cgcctcctggagatcgtg 347
 |||||
 Db 768 cgcctcctggagatcgtg 785

RESULT 9
 AAH45373/C
 ID AAH45373 standard; DNA: 1704 BP.
 XX
 AC AAH45373;
 DT 11-SEP-2001 (first entry)
 XX
 DE C. thermoaminogenes lysin biosynthetic enzyme lysA DNA.
 XX
 KW Heat-resistant; lysin biosynthesis; enzyme; coryneform;
 KW aspartate-semialdehyde dehydrogenase; lysA; ds.
 XX
 OS Corynebacterium thermoaminogenes.
 XX
 PN JP2001120270-A.
 XX
 PD 08-MAY-2001.
 XX
 PF 01-NOV-1999; 99JP-0311148.
 XX
 PR 01-NOV-1999; 99JP-0311148.
 XX
 PA (AJIN) AJINOMOTO KK.
 XX
 DR WPI: 2001-364760/38.
 DR P-PSDB: AAG64045.
 XX
 PT A heat-resistant lysin biosynthetic system enzyme gene of a high
 PT temperature-resistant coryneform microbe
 XX
 PS Example 5; Page 16-18; 27pp; Japanese.
 XX
 CC The invention relates to a gene from a high temperature-resistant
 CC coryneform microbe that encodes a heat-resistant lysin biosynthetic
 CC enzyme. The enzyme has aspartate-semialdehyde dehydrogenase activity
 CC and can be used for growing amino acid-producing microbes. The
 CC present sequence encodes an enzyme of the invention.
 XX
 SQ Sequence 1704 BP; 286 A; 613 C; 525 G; 280 T; 0 other;

Query Match
 Best Local Similarity 3.0%; Score 18; DB 22; Length 1704;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 aggccccgaggagagg 258
 |||||
 Db 693 AGGCCCGAGGAGGAGG 676

RESULT 10
 AAV38502
 ID AAV38502 standard; DNA: 1770 BP.
 XX
 AC AAV38502;

XX 21-SEP-1998 (first entry)
 DI DNA encoding human glutamate-binding protein (HGLUBP).
 XX
 DE
 XX
 KW Human; glutamate-binding protein; HGLUBP; Incyte clone 386116;
 KW treatment; prevention; diagnosis; central nervous system disorder;
 KW signalling; glutamate; ischaemia; hypoglycaemia; epilepsy; convulsion;
 KW schizophrenia; neurodegeneration; amyotrophic lateral sclerosis;
 KW lathyrism; ss.
 XX
 XX Homo sapiens.
 XX
 FH Location/Qualifiers
 FT 67..1176
 FT /*tag= a
 FT /product= HGLUBP
 FT
 XX
 XX MO5821241-A1.
 XX
 XX 22-MAY-1998.
 XX
 PF 10-NOV-1997; 57WO-US20560.
 XX
 PR 14-NOV-1996; 56US-0749289.
 XX
 XX (INCY-) INCYTE PHARM INC.
 XX
 XX Bandman O, Coleman K;
 PI
 XX WPI: 1998-297868/26.
 DR
 DR P-PSDB: AAW62612.
 XX
 XX New human glutamate-binding protein - useful for treatment,
 XX prevention and diagnosis of central nervous system diseases, e.g.
 XX Alzheimer's
 XX
 XX Claim 3; Fig 1A-E; 66pp; English.
 XX
 CC The present sequence encodes human glutamate-binding protein (HGLUBP).
 CC The protein has structural homology to the rat glutamate-binding subunit
 CC of the N-methyl-D-aspartate receptor. The nucleic acid encoding HGLUBP
 CC was first identified in Incyte clone 386116 from a thymus cDNA library.
 CC HGLUBP is useful for treatment, prevention and diagnosis of central
 CC nervous system disorders that require an increase or reduction in
 CC signalling through glutamate, e.g. ischaemia, hypoglycaemia, epilepsy,
 CC convulsions, dementia related to AIDS, schizophrenia, neurodegeneration
 CC such as Alzheimer's, Huntington's, Creutzfeldt-Jakob or Parkinson's
 CC diseases, amyotrophic lateral sclerosis or lathyrism, pain, defective
 CC motor rhythm, vasomotor tone, baroreception and long-term potentiation,
 CC muscle relaxation and sedation. Antibodies can also be used to target
 CC pharmaceuticals to HGLUBP-expressing cells or to diagnose diseases
 CC associated with HGLUBP expression and to monitor treatment.
 XX
 SQ Sequence 1770 BP; 325 A; 629 C; 448 G; 363 T; 1 other;

Query Match
 Best Local Similarity 3.0%; Score 18; DB 19; Length 1770;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 cgcctcctggagatcgtg 347
 |||||
 Db 976 cgcctcctggagatcgtg 993

RESULT 11
 AAH98783
 ID AAH98783 standard; cDNA: 1776 BP.
 XX
 AC AAH98783;
 XX
 DT 12-OCT-2001 (first entry)

1


```
DE Human TLE-2 gene.
XX TLE-2; transducin-like enhancer of split protein; cell fate;
KW differentiation; cervix cancer; breast cancer; psoriasis; baldness;
KW ss.
XX Homo sapiens.
XX Key Location/Qualifiers
FT CDS 26..2257
FT /*tag= a
XX
XX
XX W09407522-A.
XX
XX 14-APR-1994.
PD
XX 30-SEP-1993; 93WO-US09339.
PF
XX 30-SEP-1992; 92US-0954813.
PR
XX (UYVA ) UNIV YALE.
PA
XX Artavanis-tsakonas S, Stifani S;
PI
XX WPI; 1994-135221/16.
DR P-PSDB; AAR51110.
XX
XX Transducin-like enhancer or split proteins and nucleic acids-
PT are for treatment of disorders of cell fate or differentiation
PT e.g. cervical cancer, breast cancer, psoriasis, baldness etc.
XX
XX Disclosure; Page 91-94; 147pp; English.
XX
XX The nucleotide and deduced aa sequences of human transducin-like
CC enhancer of split proteins TLE-1 (AAQ45333, AAR51109), TLE-2 (AAQ45334,
CC AAR51110), TLE-3 (AAQ45335, AAR51111) and TLE-4 (AAQ45336, AAR52553)
CC were determined. The aa sequences were compared with that of
CC Drosophila E(spl) m9/10 (AAR52955). Comparison of the WD-40 domains
CC of these proteins defined the consensus residues shown in AAR52954.
CC The CcN motifs of the proteins were compared with those of SV40 T
CC antigen, human c-myc, human p53, human A-myb and dorsal protein with
CC respect to nuclear localization site, and casein-kinase and cdk2-
CC kinase phosphorylation sites (sequences AAR52556-70).
XX
XX Sequence 2271 BP; 437 A; 751 C; 672 G; 411 T; 0 other;
SQ
Query Match 3.0%; Score 18; DB 15; Length 2271;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 16 tggagctgtgtcccgatg 33
DB 405 TGGAGCTGCTGCCCGATG 388
RESULT 16
AAA07836
ID AAA07836 standard; cDNA; 7746 BP.
XX
XX AAA07836;
AC
XX 07-JUL-2000 (first entry)
DT
XX Human homologue of UNC-53 (Hs-UNC-53/2) nucleotide sequence.
DE
XX UNC-53; Caenorhabditis elegans; microtubule; neural regeneration;
KW anticancer; anti-neurodegeneration; antifibrotic; anti-adhesive; human;
KW antisclerotic; antimeastatic; anti-arthritis; autoimmune disease; ss.
XX
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH
FT
```

```
FI misc_feature 1..366
FI /*tag= a
FI /*note= "this region can be replaced with one of the
FI three sequences shown in AAA07837 to AAA07839;
FI this provides three variants for the 5' end
FI of the gene"
FI 5425..5433
FI misc_feature
FI /*tag= b
FI /*note= "this region is found to be absent in a
FI variant cDNA from Hela and colorectal
FI adenocarcinoma tissue"
FI 5924..6024
FI misc_feature
FI /*tag= c
FI /*note= "this region is found to be absent in a
FI variant cDNA isolated from frontal cortex;
FI Absence of this fragment results in an out-of-
FI frame deletion of 101 basepairs, resulting in
FI premature stop in translation"
XX
XX W095963080-A1.
XX
XX 05-DEC-1999.
XX
XX 02-JUN-1999; 99WO-EP03848.
XX
XX 03-JUN-1998; 98GB-QJ11962.
XX
XX (JANC ) JANSSEN PHARM NV.
XX
XX Luyten WHML, De Raeymaeker MC, Geysen JJGH, Bogaert TAOE;
XX Maerten LJS, Verhasselt P, Van De Craen M;
PI
XX WPI; 2000-116370/10.
XX
XX P-PSDB; AAR55565.
XX
XX Novel proteins and nucleic acids e.g. for treating neurodegeneration
XX
XX Claim 94; Fig 1c; 146pp; English.
XX
XX The invention provides vertebrate (human) protein homologue of a UNC-53
XX protein of Caenorhabditis elegans. The UNC-53 binds to microtubules or
XX their plus ends. The UNC-53 sequences are used to promote neural
XX regeneration, revascularization and wound healing; also for treating
XX neurodegenerative disease, acute traumatic injury, fibrotic disease and
XX autoimmune diseases (e.g. rheumatoid arthritis and sclerosis). The UNC-53
XX polynucleotides can be used for recombinant production of the proteins,
XX as a source of probes for detecting allelic variants and polymorphisms,
XX for sequencing genomic DNA and for detecting UNC-53 expression; and as
XX source of therapeutic antisense sequences. Cells that express the
XX protein are used to identify regulators of cell shape, growth, motility
XX and migration. They can also be used to identify proteins that are
XX involved in signal transduction pathways also involving UNC-53, and to
XX identify compounds that alter attachment of UNC-53 to microtubules. A
XX target gene coupled to a UNC-53 encoding sequence may be used to deliver
XX the target gene to a cellular microtubule or its plus ends. The present
XX sequence represents the nucleotide sequence of the second human homologue
XX of UNC-53, designated Hs-UNC-53/2.
XX
XX Sequence 7746 BP; 1563 A; 2335 C; 2010 G; 1410 T; 8 other;
SQ
Query Match 3.0%; Score 18; DB 21; Length 7746;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 284 gctggagactctgtcccg 301
DB 1386 gctggagactctgtcccg 1403
RESULT 17
AAA0583
ID AAX40983 standard; cDNA; 298 BP.
```

```

XX AC AAX40983;
XX DT 18-JUN-1999 (first entry)
XX DE Human secreted protein 5' EST SEQ ID NO: 195.
XX KW Human: secreted protein; EST: expressed sequence tag; diagnosis;
XX KW forensic; gene therapy; chromosome mapping; signal peptide;
XX KW upstream regulatory sequence; cytokine activity; cell proliferation;
XX KW differentiation; haematopoiesis regulation; tissue growth regulation;
XX KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
XX KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX OS Homo sapiens.
XX PN WO9906554-A2.
XX PD 11-FEB-1999.
XX PF 31-JUL-1998; 98WO-IB01238.
XX PR 01-AUG-1997; 97US-0905134.
XX PA (GEST ) GENSET.
XX PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX DR WPI: 1999-153784/13.
XX DR P-PSDB; AAY12150.
XX PT New nucleic acids encoding human secreted proteins - obtained from
XX PT cDNA libraries prepared from kidney, fetal kidney, dystrophic
XX PT muscle, muscle and heart tissue
XX PS Claim 1; Page 292; 622pp; English.
XX CC AAX40826 to AAX41093 represent 5' expressed sequence tags (ESTs) for
XX CC human secreted proteins, and encode the proteins given in AAY01602 and
XX CC AAY11994 to AAY12260, respectively. The proteins given represent the
XX CC signal peptide and an N-terminal fragment of a secreted protein. The
XX CC nucleic acid sequences can be used for producing secreted human gene
XX CC products. They can also be used to develop products for diagnostic and
XX CC therapy. The proteins obtained may have cytokine activity, cell
XX CC proliferation/differentiation activity, haematopoiesis regulating
XX CC activity, tissue growth regulating activity, reproductive hormone
XX CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
XX CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX CC activity, tumour inhibition activity or other activities. The products
XX CC can be used in forensic, gene therapy and chromosome mapping procedures.
XX CC The sequences can also be used for obtaining corresponding promoter
XX CC sequences. The nucleic acids encoding the signal peptide can be used
XX CC for directing extracellular secretion of a polypeptide or the insertion
XX CC of a polypeptide into a membrane, or importing a polypeptide into
XX CC a cell.
XX SQ Sequence 298 BP; 73 A; 66 C; 101 G; 58 T; 0 other;

Query Match 2.8%; Score 17; DB 20; Length 298;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 225 cgggcggctgctgaaga 241
DB 125 cgggcggctgctgaaga 141

RESULT 18
AAC06962/c
ID AAC06962 standard; cDNA: 385 BP.
XX AC AAC06962;

```

```

XX DT 06-OCT-2000 (first entry)
XX DE Human secreted protein 5' EST, SEQ ID NO: 11037.
XX KW Human; 5' EST: expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping; ss.
XX OS Homo sapiens.
XX PN EP1033401-A2.
XX PD 06-SEP-2000.
XX PF 21-FEB-2000; 2000EP-0-000610.
XX PR 26-FEB-1999; 99US-0122487.
XX PA (GEST ) GENSET.
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX DR WPI: 2000-500381/45.
XX CC New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX CC obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX CC diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX CC Claim 1; SEQ ID 11037; 71pp + CD-ROM; English.
XX CC The present sequence is one of a large number of 5' ESTs derived from
XX CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX CC identified within the present sequence. The 5' ESTs were prepared from
XX CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX CC sequences usually correspond mainly to the 3' untranslated region (UTR)
XX CC of the mRNA because they are often obtained from oligo-dT primed cDNA
XX CC libraries. Such ESTs are not well suited for isolating cDNA sequences
XX CC derived from the 5' ends of mRNAs and even in those cases where longer
XX CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX CC They are used to obtain upstream regulatory sequences and to design
XX CC expression and secretion vectors.
XX SQ Sequence 385 BP; 94 A; 83 C; 122 G; 84 T; 2 other;

Query Match 2.8%; Score 17; DB 21; Length 385;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 aaggatgcccaagagga 493
DB 185 AAGGATGCCAAGAGGA 169

RESULT 19
AAC06983/c
ID AAC06983 standard; cDNA: 413 BP.
XX AC AAC06983;
XX DT 06-OCT-2000 (first entry)
XX DE Human secreted protein 5' EST, SEQ ID NO: 11058.
XX KW Human; 5' EST: expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping; ss.
XX OS Homo sapiens.
XX PN EP1033401-A2.

```


OS	Allium porrum.	CDs	3..602	
XX		FI	/tag- a	
PN	US6060644-A.	FI	/partial	
XX		FI	/product= "Leek glossy8 homologue protein fragment"	
PD	05-MAY-2000.	FI	/note= "No initiation or termination codon given in the specification"	
XX	25-DEC-1995; 95US-0581148.	FI	/transl_except= (pos:282..285, aa:Gly)	
XX		FI	/note= "There is an apparent 1 base insertion which alters the reading frame"	
PR	24-MAR-1994; 94US-0218028.	FI	/transl_except= (pos:463..464, aa:Ala)	
XX		FI	/note= "There is an apparent 1 base deletion which alters the reading frame"	
PA	(IOWA) UNIV IOWA STATE RES FOUND INC.	FI	misc_feature 603..872	
XX	Schnable PS, Nikolau BJ, Xu X, Xia Y, Robertson DS, Hansen JD;	FI	/tag- C	
XX	WPI; 2000-349707/30.	FI	/note= "The amino acids encoded by this region are not included in the corresponding protein"	
XX	Transforming plants such as maize and canola, for producing new plant varieties having disease and pest resistance involves introducing cuticular lipid genes into the plant genome	XX		
PS	Claim 1; Column 25-26; 69pp; English.	XX	US6060644-A.	
CC	The invention relates to transforming a plant cell with either a nucleic acid encoding a plant cuticular lipid biosynthetic protein, or an antisense nucleic acid sequence targeted to a plant cuticular lipid gene, and then generating a plant from the plant cell. The cuticular lipid genes that may be used in the invention are given in	XX	05-MAY-2000.	
CC	AA57871-A57884 and AA57891. The cuticle consists of a meshwork (cutin) of cross- esterified polymerised hydroxy-fatty acids embedded in a complex mixture of nonpolar lipids (the cuticular wax). The cuticle protects and strengthens the plants, prevents evaporation of internal fluids and filters UV radiation. Mutations in the cuticular lipid genes affects the quantity and composition of cuticular lipids; in maize, 17 loci (the glossy (gl) genes) have been identified as being involved in cuticular lipid biosynthesis or control. The method is useful for introducing a nucleic acid into a cell preferably of maize, soybean, rapeseed, canola, cotton, safflower, peanut, palm or sunflower plant and generating a plant from the plant cell. The transformed plants have improved environmental resistance relating to wind, frost, UV or drought; fungal, bacterial or viral disease resistance; pest resistance; and altered lipid content in seeds. Sequences AAA57874-A57875 represent leek glossy8 (gl8) homologue partial cDNA sequences.	XX	25-DEC-1995; 95US-0581148.	
CC	Sequence 872 BP; 269 A; 219 C; 165 G; 219 T; 0 other;	XX	24-MAR-1994; 94US-0218028.	
Query Match 2.8%; Score 17; DB 21; Length 872;		XX	(IOWA) UNIV IOWA STATE RES FOUND INC.	
Best Local Similarity 100.0%; Pred. No. 93;		XX	Schnable PS, Nikolau BJ, Xu X, Xia Y, Robertson DS, Hansen JD;	
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		XX	WPI; 2000-349707/30.	
QY 475 tcaaggatgccaaagaag 491		XX	Transforming plants such as maize and canola, for producing new plant varieties having disease and pest resistance involves introducing cuticular lipid genes into the plant genome	
DB 209 TCAAGGATGCCAAGAAG 193		XX	Claim 1; Column 25-26; 69pp; English.	
RESULT 22		XX	The invention relates to transforming a plant cell with either a nucleic acid encoding a plant cuticular lipid biosynthetic protein, or an antisense nucleic acid sequence targeted to a plant cuticular lipid gene, and then generating a plant from the plant cell. The cuticular lipid genes that may be used in the invention are given in	
AAA57875		XX	AA57871-A57884 and AA57891. The cuticle consists of a meshwork (cutin) of cross- esterified polymerised hydroxy-fatty acids embedded in a complex mixture of nonpolar lipids (the cuticular wax). The cuticle protects and strengthens the plants, prevents evaporation of internal fluids and filters UV radiation. Mutations in the cuticular lipid genes affects the quantity and composition of cuticular lipids; in maize, 17 loci (the glossy (gl) genes) have been identified as being involved in cuticular lipid biosynthesis or control. The method is useful for introducing a nucleic acid into a cell preferably of maize, soybean, rapeseed, canola, cotton, safflower, peanut, palm or sunflower plant and generating a plant from the plant cell. The transformed plants have improved environmental resistance relating to wind, frost, UV or drought; fungal, bacterial or viral disease resistance; pest resistance; and altered lipid content in seeds. Sequences AAA57874-A57875 represent leek glossy8 (gl8) homologue partial cDNA sequences.	
AC AAA57875;		XX	Sequence 872 BP; 269 A; 219 C; 165 G; 219 T; 0 other;	
XX 10-OCT-2000 (first entry)		XX	Query Match 2.8%; Score 17; DB 21; Length 872;	
DT Leek glossy8 (gl8) homologue partial cDNA, SEQ ID NO:6.		XX	Best Local Similarity 100.0%; Pred. No. 93;	
DE Leek; glossy8 homologue; gl8; cuticle; cuticular lipid biosynthesis; lipid composition; lipid quantity; disease resistance; pest resistance; wind resistance; frost resistance; UV resistance; transgenic plant; antisense inhibition; ds.		XX	Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
XX Allium porrum.		XX	QY 475 tcaaggatgccaaagaag 491	
OS Allium porrum.		XX	DB 209 TCAAGGATGCCAAGAAG 193	
XX Key Location/Qualifiers		XX	RESULT 22	
PH		XX	AAA57875	
		XX	AC AAA57875;	
		XX	XX 10-OCT-2000 (first entry)	
		XX	DT Leek glossy8 (gl8) homologue partial cDNA, SEQ ID NO:6.	
		XX	DE Leek; glossy8 homologue; gl8; cuticle; cuticular lipid biosynthesis; lipid composition; lipid quantity; disease resistance; pest resistance; wind resistance; frost resistance; UV resistance; transgenic plant; antisense inhibition; ds.	
		XX	XX Allium porrum.	
		XX	OS Allium porrum.	
		XX	XX Key Location/Qualifiers	
		XX	PH	

XX	AAH50961 standard; cDNA; 991 BP.	XX	21-AUG-2000 (first entry)	XX	Human G protein-coupled receptor GPR17 cDNA.	XX	G protein-coupled receptor; GPCR; constitutively active; intracellular loop 3; transmembrane domain 6; drug screening; agonist; antagonist; ss.	XX	Homo sapiens.	XX	WO200022125-A1.	XX	20-APR-2000.	XX	12-OCT-1999; PWO-US23938.	XX	13-OCT-1998; PUS-0170456.	XX	(ARENA) ARENA PHARM INC.	XX	Behan DP, Chalters DI, Liaw CW.	XX	WPI: 2000-329167/28.	XX	P-PSDB; AAAY50616.	XX	Non-endogenous constitutively activated human G protein-coupled receptors, useful for identifying agonists for use as pharmaceutical agents.	XX	Example 1; Page 125-126; 341pp; English.	XX	The invention relates to constitutively active, non-endogenous versions of endogenous human orphan G protein-coupled receptors (GPCRs, AAAY50643- AAAY50677 and AAAY50683-Y50687), and to DNA encoding them (AAA30709-A30743 and AAA30775-A30779). The mutant proteins of the invention contain a mutation in a portion of the protein comprising the intracellular loop 3 (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, x, is substituted for an endogenous residue in TM6 at a position 16 amino acids N-terminal of an endogenous proline in TM6 to form a sequence x-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg or Ala, and is preferably Lys. When the endogenous residue at this position is Lys, this residue is replaced by His, Arg or preferably Ala. The 15 amino acid stretch between the substituted amino acid and the Pro may be endogenous, non-endogenous, or a mixture of endogenous and non-endogenous residues. The constitutively active GPCRs are useful for identifying antagonists, agonists and partial agonists for use as pharmaceutical agents. The mutant proteins are also useful in research settings for elucidating the roles of the receptors in normal and diseased conditions. Antagonists for a particular GPCR are useful for treating diseases and disorders associated with that receptor. Because the novel mutant GPCRs are constitutively active, they can be used directly for screening of compounds without the need for endogenous ligands. The present sequence represents cDNA encoding a human wild-type GPCR used in an exemplification of the invention. This was cloned and subjected to site-directed mutagenesis (SDM) to generate DNA encoding the corresponding mutant of the invention.	XX	Sequence 1020 BP; 174 A; 364 C; 261 G; 221 T; 0 other;	XX	Query Match 2.8%; Score 17; DB 21; Length 1020; Best Local Similarity 100.0%; Pred. No. 93; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	QY	323 caccacagcgcctctg 339 	DB	810 caccacagcgcctctg 826	XX	RESULT 25	XX	AAA30718	XX	ID AAA30718 standard; cDNA; 1020 BP.	XX	AAH50961 standard; cDNA; 991 BP.	XX	AAH50961;	XX	28-AUG-2001 (first entry)	XX	Phosphatidylinositol synthase partial clone nucleotide sequence.	XX	Moss; Physcomitrella patens; lipid metabolism related protein; LMRP; lipid biosynthesis; lipid modification; lipid degradation; cofactor; fatty acid transport; genetic engineering; fatty acid; enzyme; plant; microorganism; polyunsaturated fatty acid; oilseed plant; maize; wheat; biotic stress tolerance; abiotic stress tolerance; rye; oat; triticale; rice; barley; soybean; peanut; cotton; rapeseed; canola; manihot; pepper; sunflower; tagetes; potato; tobacco; eggplant; tomato; Vicia; pea; alfalfa; coffee; cacao; tea; Salix; oil palm; coconut; perennial grass; forage crop; ss.	XX	Physcomitrella patens.	XX	WO200138484-A2.	XX	31-MAY-2001.	XX	22-NOV-2000; 2000WO-EP11615.	XX	25-NOV-1999; 99WO-EP09108.	XX	(BADI) BASF PLANT SCI GMBH.	XX	Lerchl J, Renz A, Ehrhardt T, Reindl A, Cirpus P, Bischoff F; Frank M, Freund A, Duwenig E, Schmidt R, Reski R;	XX	WPI: 2001-367669/38.	XX	Nucleic acids encoding lipid metabolism related proteins from Physcomitrella patens useful to produce fine chemicals in modified organisms, particularly polyunsaturated fatty acids in oilseed plants - Claim 7; Page 105; 120pp; English.	XX	The present invention describes isolated nucleic acid sequences which encode lipid metabolism related proteins (LMRP). The LMRP nucleic acids can be used to modify lipids and fatty acids, cofactors and enzymes in microorganisms and plants, particularly to produce polyunsaturated fatty acids, and are especially useful in oilseed plants. The nucleic acids may also confer biotic or abiotic stress tolerance, particularly to maize, wheat, rye, oat, triticale, rice, barley, soybean, peanut, cotton, rapeseed, canola, manihot, pepper, sunflower, tagetes, potato, tobacco, eggplant, tomato, Vicia species, pea, alfalfa, coffee, cacao, tea, Salix species, oil palm, coconut, perennial grasses and forage crops. AAH50878 to AAH50882 represent primers used in the exemplification of the present invention. AAH50883 to AAH50968 represents LMRP nucleotide sequences, and AAG80843 to AAG80928 represent LMRP protein sequences, given in the present invention.	XX	Sequence 991 BP; 281 A; 214 C; 209 G; 287 T; 0 other;	XX	Query Match 2.8%; Score 17; DB 22; Length 991; Best Local Similarity 100.0%; Pred. No. 93; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	QY	143 gccactccgaagctgc 159 	DB	240 GCCACTCCGAAGCTGC 224	XX	RESULT 24	XX	AAA30601	XX	ID AAA30601 standard; cDNA; 1020 BP.	XX	AAH50961 standard; cDNA; 991 BP.	XX	AAH50961;	XX	28-AUG-2001 (first entry)	XX	Phosphatidylinositol synthase partial clone nucleotide sequence.	XX	Moss; Physcomitrella patens; lipid metabolism related protein; LMRP; lipid biosynthesis; lipid modification; lipid degradation; cofactor; fatty acid transport; genetic engineering; fatty acid; enzyme; plant; microorganism; polyunsaturated fatty acid; oilseed plant; maize; wheat; biotic stress tolerance; abiotic stress tolerance; rye; oat; triticale; rice; barley; soybean; peanut; cotton; rapeseed; canola; manihot; pepper; sunflower; tagetes; potato; tobacco; eggplant; tomato; Vicia; pea; alfalfa; coffee; cacao; tea; Salix; oil palm; coconut; perennial grass; forage crop; ss.	XX	Physcomitrella patens.	XX	WO200138484-A2.	XX	31-MAY-2001.	XX	22-NOV-2000; 2000WO-EP11615.	XX	25-NOV-1999; 99WO-EP09108.	XX	(BADI) BASF PLANT SCI GMBH.	XX	Lerchl J, Renz A, Ehrhardt T, Reindl A, Cirpus P, Bischoff F; Frank M, Freund A, Duwenig E, Schmidt R, Reski R;	XX	WPI: 2001-367669/38.	XX	Nucleic acids encoding lipid metabolism related proteins from Physcomitrella patens useful to produce fine chemicals in modified organisms, particularly polyunsaturated fatty acids in oilseed plants - Claim 7; Page 105; 120pp; English.	XX	The present invention describes isolated nucleic acid sequences which encode lipid metabolism related proteins (LMRP). The LMRP nucleic acids can be used to modify lipids and fatty acids, cofactors and enzymes in microorganisms and plants, particularly to produce polyunsaturated fatty acids, and are especially useful in oilseed plants. The nucleic acids may also confer biotic or abiotic stress tolerance, particularly to maize, wheat, rye, oat, triticale, rice, barley, soybean, peanut, cotton, rapeseed, canola, manihot, pepper, sunflower, tagetes, potato, tobacco, eggplant, tomato, Vicia species, pea, alfalfa, coffee, cacao, tea, Salix species, oil palm, coconut, perennial grasses and forage crops. AAH50878 to AAH50882 represent primers used in the exemplification of the present invention. AAH50883 to AAH50968 represents LMRP nucleotide sequences, and AAG80843 to AAG80928 represent LMRP protein sequences, given in the present invention.	XX	Sequence 991 BP; 281 A; 214 C; 209 G; 287 T; 0 other;	XX	Query Match 2.8%; Score 17; DB 22; Length 991; Best Local Similarity 100.0%; Pred. No. 93; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	QY	143 gccactccgaagctgc 159 	DB	240 GCCACTCCGAAGCTGC 224	XX	RESULT 24	XX	AAA30601	XX	ID AAA30601 standard; cDNA; 1020 BP.	XX	AAH50961 standard; cDNA; 991 BP.	XX	AAH50961;	XX	28-AUG-2001 (first entry)	XX	Phosphatidylinositol synthase partial clone nucleotide sequence.	XX	Moss; Physcomitrella patens; lipid metabolism related protein; LMRP; lipid biosynthesis; lipid modification; lipid degradation; cofactor; fatty acid transport; genetic engineering; fatty acid; enzyme; plant; microorganism; polyunsaturated fatty acid; oilseed plant; maize; wheat; biotic stress tolerance; abiotic stress tolerance; rye; oat; triticale; rice; barley; soybean; peanut; cotton; rapeseed; canola; manihot; pepper; sunflower; tagetes; potato; tobacco; eggplant; tomato; Vicia; pea; alfalfa; coffee; cacao; tea; Salix; oil palm; coconut; perennial grass; forage crop; ss.	XX	Physcomitrella patens.	XX	WO200138484-A2.	XX	31-MAY-2001.	XX	22-NOV-2000; 2000WO-EP11615.	XX	25-NOV-1999; 99WO-EP09108.	XX	(BADI) BASF PLANT SCI GMBH.	XX	Lerchl J, Renz A, Ehrhardt T, Reindl A, Cirpus P, Bischoff F; Frank M, Freund A, Duwenig E, Schmidt R, Reski R;	XX	WPI: 2001-367669/38.	XX	Nucleic acids encoding lipid metabolism related proteins from Physcomitrella patens useful to produce fine chemicals in modified organisms, particularly polyunsaturated fatty acids in oilseed plants - Claim 7; Page 105; 120pp; English.	XX	The present invention describes isolated nucleic acid sequences which encode lipid metabolism related proteins (LMRP). The LMRP nucleic acids can be used to modify lipids and fatty acids, cofactors and enzymes in microorganisms and plants, particularly to produce polyunsaturated fatty acids, and are especially useful in oilseed plants. The nucleic acids may also confer biotic or abiotic stress tolerance, particularly to maize, wheat, rye, oat, triticale, rice, barley, soybean, peanut, cotton, rapeseed, canola, manihot, pepper, sunflower, tagetes, potato, tobacco, eggplant, tomato, Vicia species, pea, alfalfa, coffee, cacao, tea, Salix species, oil palm, coconut, perennial grasses and forage crops. AAH50878 to AAH50882 represent primers used in the exemplification of the present invention. AAH50883 to AAH50968 represents LMRP nucleotide sequences, and AAG80843 to AAG80928 represent LMRP protein sequences, given in the present invention.	XX	Sequence 991 BP; 281 A; 214 C; 209 G; 287 T; 0 other;	XX	Query Match 2.8%; Score 17; DB 22; Length 991; Best Local Similarity 100.0%; Pred. No. 93; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	QY
----	----------------------------------	----	---------------------------	----	--	----	---	----	---------------	----	-----------------	----	--------------	----	---------------------------	----	---------------------------	----	--------------------------	----	---------------------------------	----	----------------------	----	--------------------	----	--	----	--	----	--	----	--	----	---	----	------------------------------	----	--------------------------	----	-----------	----	----------	----	--------------------------------------	----	----------------------------------	----	-----------	----	---------------------------	----	--	----	---	----	------------------------	----	-----------------	----	--------------	----	------------------------------	----	----------------------------	----	-----------------------------	----	---	----	----------------------	----	---	----	---	----	---	----	--	----	------------------------------	----	--------------------------	----	-----------	----	----------	----	--------------------------------------	----	----------------------------------	----	-----------	----	---------------------------	----	--	----	---	----	------------------------	----	-----------------	----	--------------	----	------------------------------	----	----------------------------	----	-----------------------------	----	---	----	----------------------	----	---	----	---	----	---	----	--	----	------------------------------	----	--------------------------	----	-----------	----	----------	----	--------------------------------------	----	----------------------------------	----	-----------	----	---------------------------	----	--	----	---	----	------------------------	----	-----------------	----	--------------	----	------------------------------	----	----------------------------	----	-----------------------------	----	---	----	----------------------	----	---	----	---	----	---	----	--	----

XX	AAF88323 standard; DNA: 1194 BP.	XX	AAF88323 standard; DNA: 1194 BP.
AC	AAF88323;	AC	AAF88323;
XX	28-AUG-2001 (first entry)	XX	28-AUG-2001 (first entry)
XX	S. spinosa DNA fragment encoding ORF5, SEQ ID 17.	XX	S. spinosa DNA fragment encoding ORF5, SEQ ID 17.
XX	Forosamine; trimethylrhannose; polyketide synthase; biosynthesis;	XX	Forosamine; trimethylrhannose; polyketide synthase; biosynthesis;
KW	spinosyn; polyketide aglycone; transgenic plant; insect resistance;	KW	spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW	macrolide; insecticidal; O-methyltransferase; ds.	KW	macrolide; insecticidal; O-methyltransferase; ds.
XX	Saccharopolyspora spinosa.	XX	Saccharopolyspora spinosa.
OS	DEL9557268-A1.	OS	DEL9557268-A1.
XX	08-MAR-2001.	XX	08-MAR-2001.
XX	29-NOV-1999; 99DE-1057268.	XX	29-NOV-1999; 99DE-1057268.
XX	27-AUG-1999; 99DE-1040596.	XX	27-AUG-1999; 99DE-1040596.
XX	(FARB) BAYER AG.	XX	(FARB) BAYER AG.
XX	Eberz G, Moehrl V, Froede R, Velten R, Salas JA;	XX	Eberz G, Moehrl V, Froede R, Velten R, Salas JA;
PI	WPI: 2001-267102/28.	PI	WPI: 2001-267102/28.
DR	P-PSDB; AAB70553.	DR	P-PSDB; AAB70553.
XX	New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for	XX	New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT	recombinant production of insecticidal spinosyns and their derivatives	PT	recombinant production of insecticidal spinosyns and their derivatives
PT	Claim 7a; Page 121-124; 354pp; German.	PT	Claim 7a; Page 121-124; 354pp; German.
PS	This invention describes a novel method nucleic acid (I) and its encoded	PS	This invention describes a novel method nucleic acid (I) and its encoded
XX	polypeptide (II) containing at least one region that encodes an enzymatic	XX	polypeptide (II) containing at least one region that encodes an enzymatic
CC	activity involved in biosynthesis of spinosyns. (I) are used (i) to	CC	activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC	identify, inactivate or modulate genes involved in the biosynthesis of	CC	identify, inactivate or modulate genes involved in the biosynthesis of
CC	(ii); (iii) to generate a library of polyketide synthases; (iii) for	CC	(ii); (iii) to generate a library of polyketide synthases; (iii) for
CC	adding forosamine or trimethylrhannose to a spinosyn or polyketide	CC	adding forosamine or trimethylrhannose to a spinosyn or polyketide
CC	aglycone; and (iv) for recombinant production of the precursors or	CC	aglycone; and (iv) for recombinant production of the precursors or
CC	enzymes, which are used for production of insects. (i) are also useful as	CC	enzymes, which are used for production of insects. (i) are also useful as
CC	derivatives, including production of transgenic plants that express (ii)	CC	derivatives, including production of transgenic plants that express (ii)
CC	and thus have increased resistance to insects. (i) are also useful as	CC	and thus have increased resistance to insects. (i) are also useful as
CC	markers for sequencing of the Saccharopolyspora spinosa genome. (ii) are	CC	markers for sequencing of the Saccharopolyspora spinosa genome. (ii) are
CC	also be used to raise specific antibodies, useful for identifying	CC	also be used to raise specific antibodies, useful for identifying
CC	expression clones in a gene bank. Cells transformed with (i) may produce	CC	expression clones in a gene bank. Cells transformed with (i) may produce
CC	(ii) at significantly increased levels or produce new derivatives of	CC	(ii) at significantly increased levels or produce new derivatives of
CC	(ii). This sequence, ORF 5, encodes an S. spinosa O-methyltransferase	CC	(ii). This sequence, ORF 5, encodes an S. spinosa O-methyltransferase
XX	protein.	XX	protein.
XX	Sequence 1194 BP; 230 A; 408 C; 361 G; 195 T; 0 other;	XX	Sequence 1194 BP; 230 A; 408 C; 361 G; 195 T; 0 other;
SY	Query Match 2.8%; Score 17; DB 22; Length 1194;	SY	Query Match 2.8%; Score 17; DB 22; Length 1194;
	Best Local Similarity 100.0%; Pred. No. 92;		Best Local Similarity 100.0%; Pred. No. 92;
	Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	329 ggcacatccctggagatcg 345	QY	329 ggcacatccctggagatcg 345
DB		DB	
DB	579 ggcacatccctggagatcg 595	DB	579 ggcacatccctggagatcg 595
RESULT 27		RESULT 27	
AAC93490		AAC93490	
ID	AAC93490 standard; cDNA; 1258 BP.	ID	AAC93490 standard; cDNA; 1258 BP.
XX	AAC93490;	XX	AAC93490;
AC	16-FEB-2001 (first entry)	AC	16-FEB-2001 (first entry)
XX		XX	
DT		DT	
XX		XX	

XX	AA30718;	XX	AA30718;
AC	21-AUG-2000 (first entry)	AC	21-AUG-2000 (first entry)
XX	DNA encoding human mutant G protein-coupled receptor GPR17 (V234K).	XX	DNA encoding human mutant G protein-coupled receptor GPR17 (V234K).
XX	G protein-coupled receptor; GPCR; constitutively active;	XX	G protein-coupled receptor; GPCR; constitutively active;
KW	intracellular loop 3; transmembrane domain 6; drug screening;	KW	intracellular loop 3; transmembrane domain 6; drug screening;
KW	agonist; antagonist; mutant; ss.	KW	agonist; antagonist; mutant; ss.
XX	Homo sapiens.	XX	Homo sapiens.
OS	Synthetic.	OS	Synthetic.
XX	WO200022129-A1.	XX	WO200022129-A1.
PN	20-APR-2000.	PN	20-APR-2000.
PD	12-OCT-1999; 99WO-US23938.	PD	12-OCT-1999; 99WO-US23938.
XX	13-OCT-1998; 98US-0170496.	XX	13-OCT-1998; 98US-0170496.
XX	(AREN-) ARENA PHARM INC.	XX	(AREN-) ARENA PHARM INC.
XX	Behan DP, Chalmers DT, Liaw CW;	XX	Behan DP, Chalmers DT, Liaw CW;
PI	WPI: 2000-329165/28.	PI	WPI: 2000-329165/28.
DR	P-PSDB; AA30652.	DR	P-PSDB; AA30652.
XX	Non-endogenous constitutively activated human G protein-coupled	XX	Non-endogenous constitutively activated human G protein-coupled
PT	receptors, useful for identifying agonists for use as pharmaceutical	PT	receptors, useful for identifying agonists for use as pharmaceutical
PT	agents	PT	agents
XX	Example 2; Page 232-233; 341pp; English.	XX	Example 2; Page 232-233; 341pp; English.
PS	The invention relates to constitutively active, non-endogenous versions	PS	The invention relates to constitutively active, non-endogenous versions
XX	of endogenous human orphan G protein-coupled receptors (GPCRs, AA30643-	XX	of endogenous human orphan G protein-coupled receptors (GPCRs, AA30643-
CC	AA30677 and AA30683-Y90687), and to DNA encoding them (AA30709-A30713	CC	AA30677 and AA30683-Y90687), and to DNA encoding them (AA30709-A30713
CC	and AA30775-A30779). The mutant proteins of the invention contain a	CC	and AA30775-A30779). The mutant proteins of the invention contain a
CC	mutation in a portion of the protein comprising intracellular loop 3	CC	mutation in a portion of the protein comprising intracellular loop 3
CC	(IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,	CC	(IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC	is substituted for an endogenous residue in IC3 at a position 16 amino	CC	is substituted for an endogenous residue in IC3 at a position 16 amino
CC	acids N-terminal of an endogenous proline in TM6 to form a sequence	CC	acids N-terminal of an endogenous proline in TM6 to form a sequence
CC	X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg	CC	X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg
CC	or Ala, and is preferably Lys. When the endogenous residue at this	CC	or Ala, and is preferably Lys. When the endogenous residue at this
CC	position is Lys, this residue is replaced by His, Arg or preferably Ala.	CC	position is Lys, this residue is replaced by His, Arg or preferably Ala.
CC	The 15 amino acid stretch between the substituted amino acid and the Pro	CC	The 15 amino acid stretch between the substituted amino acid and the Pro
CC	may be endogenous, non-endogenous, or a mixture of endogenous and	CC	may be endogenous, non-endogenous, or a mixture of endogenous and
CC	non-endogenous residues. The constitutively active GPCRs are useful for	CC	non-endogenous residues. The constitutively active GPCRs are useful for
CC	identifying antagonists, agonists and partial agonists for use as	CC	identifying antagonists, agonists and partial agonists for use as
CC	pharmaceutical agents. The mutant proteins are also useful in research	CC	pharmaceutical agents. The mutant proteins are also useful in research
CC	settings for elucidating the roles of the receptors in normal and	CC	settings for elucidating the roles of the receptors in normal and
CC	diseased conditions. Antagonists for a particular GPCR are useful for	CC	diseased conditions. Antagonists for a particular GPCR are useful for
CC	treating diseases and disorders associated with that receptor. Because	CC	treating diseases and disorders associated with that receptor. Because
CC	the novel mutant GPCRs are constitutively active, they can be used	CC	the novel mutant GPCRs are constitutively active, they can be used
CC	directly for screening of compounds without the need for endogenous	CC	directly for screening of compounds without the need for endogenous
CC	ligands. Sequences AA30709- AA30743 and AA30775-A30779 represent DNAs	CC	ligands. Sequences AA30709- AA30743 and AA30775-A30779 represent DNAs
CC	encoding the mutant human GPCRs of the invention.	CC	encoding the mutant human GPCRs of the invention.
XX	Sequence 1020 BP; 177 A; 364 C; 259 G; 220 T; 0 other;	XX	Sequence 1020 BP; 177 A; 364 C; 259 G; 220 T; 0 other;
SY	Query Match 2.8%; Score 17; DB 21; Length 1020;	SY	Query Match 2.8%; Score 17; DB 21; Length 1020;
	Best Local Similarity 100.0%; Pred. No. 93;		Best Local Similarity 100.0%; Pred. No. 93;
	Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	323 caccacagcgcacatccctgg 339	QY	323 caccacagcgcacatccctgg 339
DB		DB	
DB	810 caccacagcgcacatccctgg 826	DB	810 caccacagcgcacatccctgg 826
RESULT 26		RESULT 26	
AAF88323		AAF88323	

DE Human secreted protein gene 12 SEQ ID NO:22.

XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;

KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;

KW neuroprotective; antibacterial; virucide; fungicide;

KW ophthalmological; vulnary; autoimmune disease; rheumatoid arthritis;

KW hyperproliferative disorders; cancer; cardiovascular disorder;

KW cardiac arrest; cerebrovascular disorder; nervous system disorder;

KW Alzheimer's disease; ocular disorder; wound healing; skin aging; ss.

XX Homo sapiens.

OS

PN WO2000061626-A1.

XX

PD 19-OCT-2000.

XX

PF 06-APR-2000; 2000WO-US09066.

XX

PR 09-APR-1999; 99US-0128698.

PR 20-JAN-2000; 2000US-0176926.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

XX

PI Rosen CA, Ruben SM, Komatsoulis G;

XX

XX WPI: 2000-619227/59.

DR P-PSDB; AAB51838.

DR

XX New nucleic acid molecules encoding 49 human secreted proteins for

PT diagnosing, preventing or ameliorating medical conditions and used for

PT food additives or preservatives -

XX

PS Claim 1; Page 440; 516pp; English.

XX

XX Polynucleotide sequences AAC93479 - AAC93527 represent cDNA encoding

CC human secreted proteins AAB51827 - AAB51875. Sequences AAB51876 -

CC AAB51927 represent alternative polypeptides encoded by the genes, and

CC amino acid sequences with which they share homology. The genes and

CC proteins have activities dependent on the tissues and cells in which they

CC are expressed. Examples of their activities include immunosuppressive;

CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;

CC vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial;

CC virucide; fungicide; ophthalmological; and vulnary. The secreted

CC proteins, polynucleotides, antagonists and agonists may be useful in

CC treating, preventing and/or diagnosing diseases and disorders such as

CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative

CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders

CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,

CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,

CC infections caused by bacteria, viruses and fungi and ocular disorders

CC e.g. corneal infection. The polypeptides can also be used to aid wound

CC healing and epithelial cell proliferation, to prevent skin aging due to

CC sunburn, to maintain organs before transplantation, for supporting cell

CC culture of primary tissues, to regenerate tissues and in chemotaxis. The

CC polypeptides can also be used as a food additive or preservative to

CC increase or decrease storage capabilities, fat content, lipid, protein,

CC carbohydrate, vitamins, minerals, cofactors and other nutritional

CC components. Oligonucleotides AAC93470 - AAC93478 and peptide AAB51826 are

CC used in the isolation and characterisation of the proteins and

CC polynucleotides of the invention.

XX

XX Sequence 1258 BP; 291 A; 313 C; 330 G; 306 T; 18 other;

RESULT 28

AA066178

ID AA066178 standard; DNA; 1901 BP.

XX

AC AA066178;

XX

DI 07-FEB-1995 (first entry)

XX

DE Seven transmembrane receptor (R12) coding sequence.

XX

KW Primer; seven transmembrane receptor; receptor; amplification; ss.

XX

OS Homo sapiens.

XX

PH Key Location/Qualifiers

FT CUS 701..1720

FT /*tag= a

FT /product= Seven transmembrane receptor (R12).

XX

PN W05412635-A.

XX

XX 09-JUN-1994.

XX

XX 17-NOV-1993; 93WO-US11153.

XX

XX 17-NOV-1992; 92US-0977452.

XX

XX (ICOS-) ICOS CORP.

XX

XX Godiska R, Gray PW, Schweickart VL;

XX

XX WPI: 1994-200264/24.

DR P-PSDB; AAR53753.

XX

XX DNA encoding seven transmembrane receptors - used to develop

PT prods. for use as therapeutic or diagnostic agents for conditions

PT involving the receptors.

XX

PS Example 10; Page 79-80; 100pp; English.

XX

XX PCR using two primers (AA066174, AA066175) was performed to amplify a

CC partial sequence of the R20 seven transmembrane coding sequence

CC (AA066176) from a human placenta DNA library. During the isolation

CC of the R20 gene two weakly hybridising sequences were identified

CC which had significant homology to other seven transmembrane

CC receptors. The probe was used to screen a human genomic foetal

CC liver DNA library, and while the R20 gene could not be identified in

CC this library, several weakly hybridising clones were plaque

CC purified, subcloned and sequenced. The two clones were designated

CC R2 (AA066177) and R12. This is the coding sequence of the R12 clone.

XX

XX Sequence 1901 BP; 361 A; 640 C; 508 G; 372 T; 0 other;

XX

Query Match 2.8%; Score 17; DB 15; Length 1901;

Best Local Similarity 100.0%; Pred. No. 91;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 323 caccagcgcacccctgg 339

DB 1510 caccagcgcacccctgg 1526

RESULT 29

AAV18356

ID AAV18356 standard; DNA; 1901 BP.

XX

AC AAV18356;

XX

DI 25-SEP-1998 (first entry)

XX

DR WPI: 1986-340201/52.

XX
XX
T
:
:
:

○ 9

DR P-PSDB: AAP60130.

XX New human interleukin-1 polypeptide(s) - prepd. by recombinant

PT DNA procedures for use in stimulating host immune systems, to

PT promote wound healing etc.

XX Disclosure: Fig 1: 37pp: English.

XX Murine IL-1 encoding cDNA was used to isolate the structure of a

CC partial genomic clone of human IL-1. IL-1 has utility in stimulating

CC the immune system, especially for the promotion of wound healing, as

CC a vaccine adjuvant, and in protection against neoplastic disease.

CC IL-1 is also useful in treating critically ill protein-malnourished

CC patients and in diagnostic assay systems.

XX Sequence 1985 BP: 624 A; 436 C; 364 G; 561 T; 0 other;

SQ

Query Match 2.8%; Score 17; DB 7; Length 1985;

Best Local Similarity 100.0%; Pred. No. 91;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 265 ataacgacgtgatgtca 281

Db 786 ATAAGCAGCTGATGTCA 770

|||||

RESULT 32

AAH14321/c

ID AAH14321 standard; cDNA: 2040 BP.

XX AC AAH14321;

DT 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:11685.

DE Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

KW Homo sapiens.

OS

XX EP1074617-A2.

PN 07-FEB-2001.

PD

XX 28-JUL-2000; 2000EP-0116126.

PF

XX 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELIX) HELIX RES INST.

PA

XX Ora T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI: 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

XX Claim 8; SEQ ID 11685; 2537pp + CD ROM; English.

PS

XX The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX

SQ Sequence 2040 BP: 629 A; 474 C; 468 G; 469 T; 0 other;

Query Match 2.8%; Score 17; DB 22; Length 2040;

Best Local Similarity 100.0%; Pred. No. 91;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 401 ggaagcttcagaagaaa 417

Db 1160 GGAGCTTCAGAGGAAA 1144

|||||

RESULT 33

AAH144052

ID AAT44052 standard; cDNA: 2453 BP.

XX AC AAT44052;

DT 28-FEB-1997 (first entry)

XX Human G-protein thrombin-like receptor cDNA.

DE G-protein; thrombin; receptor; diagnosis; gene therapy;

KW haemophilia; wound healing; restenosis; angina; inflammation; ss.

XX Homo sapiens.

OS

XX Key Location/Qualifiers

FT CDS 548..1567

FT /*tag= a

FT primer_bind complement (548..565)

FT /*tag= b

FT /*note= "5', primer binding site"

FT primer_bind 1547..1564

FT /*tag= C

FT /*note= "3', primer binding site"

XX WO9639438-A1.

PN

XX 12-DEC-1996.

ED

XX 06-JUN-1995; 95WO-US07180.

PF

XX 06-JUN-1995; 95WO-US07180.

PR (HUMA-) HUMAN GENOME SCI INC.

PA

XX Gocayne JD, Li Y, Ruben SM;

PI WPI: 1997-043073/04.

DR P-PSDB: AAW07617.

XX Human G-protein thrombin receptor, HIBEB69 - useful to identify

PT (ant)agonists, for treatment of angina, restenosis, wound healing

PT etc.

XX

PS Claim 1; Fig 1A-D; 58pp; English.

XX A cDNA clone (AA144092) codes for a human 7-transmembrane receptor
 CC (AAW07617) putatively identified as a G-protein thrombin receptor. It
 CC was discovered in a human infant brain cDNA library. G-protein
 CC thrombin-like receptor polynucleotides can be used to produce
 CC recombinant receptor in host cells (e.g. E. coli, COS-7, Sf9), to
 CC design probes for detecting diseases related to receptor gene
 CC mutation, and in gene therapy to treat haemophilia or promote wound
 CC healing; antisense sequences can be used to treat inflammation,
 CC restenosis, angina, etc.

XX SQ Sequence 2453 BP; 494 A; 815 C; 640 G; 504 T; 0 other;

Query Match 2.8%; Score 17; DB 18; Length 2453;

Best Local Similarity 100.0%; Pred. No. 91; Mismatches 0; Indels 0; Gaps 0;

QY 323 caccacgcgcctcctgg 339

Db 1357 caccacgcgcctcctgg 1373

RESULT 34

ID AAH66941 standard; DNA: 2766 BP.

AC AAH66941;

DT 26-SEP-2001 (first entry)

DE C glutamicum coding sequence fragment SEQ ID NO: 1976.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis; ds.

XX Corynebacterium glutamicum.

OS EPI106790-A2.

PN 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

PF 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280986.

XX (KYOW) KYOWA HAKKO KOGYO KK.

PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI: 2001-376931/40.

DR P-PSDB; AAG91722.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene

PT

PS Claim 8; SEQ ID NO: 1976; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein

CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a

CC mutant of coryneform bacterium, measuring expression amount and

CC analysing the expression profile or expression pattern of a gene derived

CC from Coryneform bacterium, and identifying a homologue of a gene derived

CC from coryneform bacterium. Coryneform bacteria are useful for producing

CC amino acids, nucleic acids, vitamins, saccharides and organic acids,

CC particularly L-lysine. The present sequence is a nucleic acid described

CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

XX SQ Sequence 2766 BP; 695 A; 610 C; 838 G; 623 T; 0 other;

Query Match 2.8%; Score 17; DB 22; Length 2766;

Best Local Similarity 100.0%; Pred. No. 90;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 ggtcagtgaggagagc 116

Db 1667 ggtcagtgaggagagc 1703

RESULT 35

ID AAC75611 standard; cDNA; 2813 BP.

XX AAC75611;

XX 06-FEB-2001 (first entry)

DE Human ORFX ORF1166 polynucleotide sequence SEQ ID NO:2331.

XX Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;

KW vulnary; antipapillary; antipapillary; antipapillary; antipapillary;

KW anticonvulsant; osteopathic; antipapillary; antipapillary; antipapillary;

KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;

KW hypotensive; dermatological; immunosuppressive; antineoplastic;

KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;

KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;

KW neurodegenerative disorder; osteoarthritis; graft vs host disease;

KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;

KW cholesterol ester storage; systemic lupus erythematosus; infection;

KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;

KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

KW bone damage; cartilage damage; antiinflammatory disease; coagulation;

KW thrombosis; contraceptive; ss.

XX Homo sapiens.

XX W:2000058473-A2.

XX 05-JUL-2000.

XX 31-MAR-2000; 2000WG-US06621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

XX 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0540763.

XX (CUKA-) CURAGEN CORP.

XX Shinketsu RA, Leach M;

XX WPI: 2000-602362/57.

XX P-PSDB; AAB41402.

XX Novel nucleic acids and peptides derived from open reading frame X,

PT useful for treating e.g. cancers, proliferative disorders,

PT neurodegenerative disorders and cardiovascular disease -

PS Claim 5; Page 1704-1706; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,

CC which represent the human ORFX open reading frames 1 to 3161. The ORFX

CC sequences have activities such as: cytostatic; hepatotropic; vulnary;

CC antipapillary; antipapillary; antipapillary; antipapillary;

CC osteopathic; anticonvulsant; antipapillary; antipapillary;

CC immunosuppressant;

CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 2813 BP; 847 A; 657 C; 637 G; 670 T; 2 other;

Query Match 2.8%; Score 17; DB 21; Length 2813;
 Best Local Similarity 100.0%; Pred. No. 90;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 401 ggagcttcagaagaaa 417
 |||||
 DB 1919 GGAGCTTCAGAGGAAA 1903

RESULT 36
 AAI60348/C
 ID AAI60348 standard; cDNA; 3092 BP.

XX AAI60348;

XX 22-OCT-2001 (first entry)
 XX Human polynucleotide SEQ ID NO 4337.

XX Human; neutropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.

XX Homo sapiens.
 XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 03-AUG-2000; 2000US-0620312.

PR 14-SEP-2000; 2000US-0653450.

PR 19-OCT-2000; 2000US-0662191.

PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

DR P-PSDB; AAM41192.

XX Novel nucleic acids and polypeptides, useful for treating disorders

PI such as central nervous system injuries -
 XX Claim 1; SEQ ID NO 4337; 10078pp; English.
 XX The invention relates to human nucleic acids (AA157758-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with neutropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX Sequence 3092 BP; 921 A; 727 C; 698 G; 746 T; 0 other;

Query Match 2.8%; Score 17; DB 22; Length 3092;
 Best Local Similarity 100.0%; Pred. No. 90;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 401 ggagcttcagaagaaa 417
 |||||
 DB 2182 GGAGCTTCAGAGGAAA 2166

RESULT 37
 AAH90050/C

ID AAH90050 standard; cDNA; 3092 BP.

XX AAH90050;

XX 01-OCT-2001 (first entry)

XX Human bone marrow cDNA, SEQ ID NO: 294.

XX Human; bone marrow; antiinflammatory; cytostatic; neuroprotective;
 KW antiviral; antibacterial; antifungal; anti-HIV; haemostatic;
 KW immunosuppressive; gene therapy; cytokine cell proliferation;
 KW cell differentiation modulator; immune disorder; infection; cancer;
 KW human immunodeficiency virus; HIV; autoimmune disorder; haemophilia; ss.

XX Homo sapiens.

XX WO200153453-A2.

XX 26-JUL-2001.

XX 23-DEC-2000; 2000WO-US34560.

XX 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 30-NOV-2000; 2000US-07250583.

XX (HYSE-) HYSEQ INC.

XX Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;
 PI Ren F, Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Drmanac RT;

XX WPI; 2001-488707/53.

DR P-PSDB: AAM00931.
 XX Novel bone-marrow-expressed polynucleotides and polypeptides, useful
 PT for treating e.g. cancer and immune deficiency disorders -
 XX
 XX Claim 1; Page 405; 648pp; English.
 PS
 XX The present sequence is one of 251 novel human polynucleotides
 CC expressed in the bone marrow. The polynucleotide and the
 CC polypeptide encoded by it are useful in the treatment of various
 CC immune deficiencies and disorders. The deficiencies and disorders may
 CC be genetic, or may be caused by a viral (e.g. HIV), bacterial or fungal
 CC infection, or may result from an autoimmune disorder, a coagulation
 CC disorder (e.g. haemophilia), inhibition of tumour cell proliferation,
 CC suppression of an inflammatory response or treatment of a nervous
 CC system disorder such as Alzheimer's disease. Detection of the presence
 CC or increased expression of the polynucleotide or the protein it
 CC encodes is useful for the diagnosis and/or prognosis of one
 CC or more types of cancer. The polynucleotide and polypeptide can be
 CC used as nutritional sources or supplements and in the screening of
 CC chemical compounds as potential drugs.
 XX Sequence 3092 BP; 921 A; 727 C; 698 G; 746 T; 0 other;
 SQ

Query Match 2.8%; Score 17; DB 22; Length 3092;
 Best Local Similarity 100.0%; Pred. No. 90;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 401 ggagcttcagaaggaaa 417
 |||||
 Db 2182 GGAGCTTCAGAGGAAA 2166

RESULT 38
 AAI58562/C
 ID AAI58562 standard; cDNA: 3382 BP.
 XX
 AC. AAI58562;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 765.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.
 DR P-PSDB: AAM39406.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 XX Claim 1; SEQ ID NO 765; 10078pp; English.
 PS
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC Specification.
 XX
 SQ Sequence 3382 BP; 573 A; 797 C; 763 G; 849 T; 0 other;

Query Match 2.8%; Score 17; DB 22; Length 3382;
 Best Local Similarity 100.0%; Pred. No. 90;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 401 ggagcttcagaaggaaa 417
 |||||
 Db 2450 GGAGCTTCAGAGGAAA 2474

RESULT 39
 AAH89937/C
 ID AAH89937 standard; cDNA: 3405 BP.
 XX
 AC. AAH89937;
 XX
 DT 01-OCT-2001 (first entry)
 XX
 DE Human bone marrow cDNA, SEQ ID NO: 68.
 XX
 KW Human; bone marrow; antiinflammatory; cytostatic; neuroprotective;
 KW antiviral; antibacterial; antifungal; anti-HIV; haemostatic;
 KW immunosuppressive; gene therapy; cytokine cell proliferation;
 KW cell differentiation modulator; immune disorder; infection; cancer;
 KW human immunodeficiency virus; HIV; autoimmune disorder; haemophilia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153453-A2.
 XX
 PD 26-JUL-2001.
 XX
 PF 23-DEC-2000; 2000WO-US34960.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 30-NOV-2000; 2000US-0250583.
 XX
 PA (HYSE-) HYSEQ INC.

PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;
PI Ren F, Wang J, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Drmanac RT;
XX WPI: 2001-488707/53.
DR P-PSDB: AAM00818.
XX Novel bone-marrow-expressed polynucleotides and polypeptides, useful
PT for treating e.g. cancer and immune deficiency disorders.
PI Claim 1: Page 252-253; 648pp; English.
XX The present sequence is one of 251 novel human polynucleotides
CC expressed in the bone marrow. The polynucleotide and the
CC polypeptide encoded by it are useful in the treatment of various
CC immune deficiencies and disorders. The deficiencies and disorders may
CC be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal
CC infection, or may result from an autoimmune disorder, a coagulation
CC disorder (e.g. haemophilia), inhibition of tumour cell proliferation,
CC suppression of an inflammatory response or treatment of a nervous
CC system disorder such as Alzheimer's disease. Detection of the presence
CC or increased expression of the polynucleotide or the protein it
CC encodes is useful for the diagnosis and/or prognosis of one
CC or more types of cancer. The polynucleotide and polypeptide can be
CC used as nutritional sources or supplements and in the screening of
CC chemical compounds as potential drugs.
XX
SQ Sequence 3405 BP; 977 A; 803 C; 770 G; 855 T; 0 other;

Query Match 2.8%; Score 17; DB 22; Length 3405;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 401 ggagcttcagaagga 417
|||||
Db 2469 GGAGCTTCAGAGGAAA 2453

RESULT 40
AAC76360/c
ID AAC76360 standard; cDNA; 5713 BP.
XX
AC AAC76360;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1915 polynucleotide sequence SEQ ID NO:3829.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antiparasitic; antiparkinsonian; nootropic; neuroprotective;
KW immunostimulant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
FA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI: 2000-502562/57.
DR P-PSUB: AAB42151.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
PS Claim 5: Page 2982-2985; 5507pp; English.
XX
CC AAC7446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antiparasitic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC osteopath; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 5713 BP; 1237 A; 1565 C; 1581 G; 1329 T; 1 other;

Query Match 2.8%; Score 17; DB 21; Length 5713;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 ccagagatgtgcccgg 314
|||||
Db 3941 CCAGAGATGTGTCCGG 3925

RESULT 41
AAH57490/c
ID AAH57490 standard; cDNA; 5813 BP.
XX
AC AAH57490;
XX
DT 10-SEP-2001 (first entry)
XX
DE Human liver cell specific cDNA sequence SEQ ID NO:330.
XX
KW Human; tissue specific; diagnosis; brain; heart; skeletal muscle;
KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;
KW metabolic disease; developmental disease; cytostatic; immunomodulatory;
KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.
XX
OS Homo sapiens.
XX
PN WO200132527-A2.
XX
PD 10-MAY-2001.
XX
XX 02-NOV-2000; 2000WO-US30396.

```

XX 04-NOV-1999; 99US-0163508.
XX (INCY-) INCYTE GENOMICS INC.
XX Sornasse T, Seilhamer JJ, Watson GA;
XX WPI; 2001-291057/30.
XX
XX New cell and tissue specific polynucleotides useful for diagnosis,
XX prognosis or monitoring of treatments for disorders where the gene is
XX associated with a cancer, immunopathology or neuropathology -
XX
XX Claim 1; Page 248-249; 327pp; English.
XX
XX AAH57161 to AAH57576 represent cell and tissue specific polynucleotide
XX sequences (I). (I) can have cytostatic, immunomodulatory and
XX neuroprotective activities, and can be used in gene therapy. (I) and
XX proteins (II) encoded by then are used in high throughput screening
XX assays to select DNA molecules, RNA molecules, peptide nucleic acids,
XX mimetics, peptides, proteins, agonists, antagonists, antibodies or
XX their fragments, immunoglobulins, inhibitors, drug compounds and
XX pharmaceutical agents. Expression of (I) in a sample indicates the
XX differentiation of embryonic stem cells into a tissue selected from
XX brain, heart, kidney, liver, lung, skeletal muscle or pancreatic
XX tissues. (I) and (II) are used to produce an expression profile that
XX defines a metabolic or developmental process, treatment, condition,
XX disease or disorder. The gene profile can be used for diagnosis,
XX prognosis or monitoring of treatments and for investigating a
XX predisposition to a disorder where the gene is associated with a
XX cancer, immunopathology or neuropathology.
XX
XX Sequence 5813 BP; 1290 A; 1665 C; 1665 G; 1193 T; 0 other:
XX
XX Query Match 2.8%; Score 17; DB 22; Length 5813;
XX Best Local Similarity 100.0%; Pred. No. 89;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 237 gaagaggccgcagga 253
DB 5426 GAAGAGGCCGCGAGAA 5410
XX
RESULT 42
AAV52730
ID AAV52730 standard; DNA; 6441 BP.
XX
XX AAV52730;
XX
XX 21-DEC-1998 (first entry)
XX
XX Human hepatocyte nuclear factor 1 beta gene (TCF2).
XX
XX Hepatocyte nuclear factor 1 beta; HNF-1 beta; MODY4; human;
XX transcription factor; maturity onset diabetes of the young;
XX diabetes; NIDDM; diagnosis; therapy; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 531..5556
XX /tag= a
XX /note= "contains introns"
XX exon 356..874
XX /tag= b
XX /number= 1
XX intron 875..1312
XX /tag= c
XX /number= 1
XX misc_difference 1100
XX /tag= d
XX /note= "n at position 1100 represents an intronic

```

```

FI exon 1313..1512
FI /tag= e
FI /number= 2
FI intron 1513..2108
FI /tag= f
FI /number= 2
FI misc_difference 1750
FI /tag= g
FI /note= "n at position 1750 represents an intronic
FI exon 2109..2373
FI /tag= h
FI /number= 3
FI intron 2374..2592
FI /tag= i
FI /number= 3
FI misc_difference 2512
FI /tag= j
FI /note= "n at position 2512 represents an intronic
FI exon 2593..2828
FI /tag= k
FI /number= 4
FI intron 2829..3266
FI /tag= l
FI /number= 4
FI misc_difference 3003
FI /tag= m
FI /note= "n at position 3003 represents an intronic
FI exon 3267..3427
FI /tag= n
FI /number= 5
FI intron 3428..3790
FI /tag= o
FI /number= 5
FI misc_difference 3609
FI /tag= p
FI /note= "n at position 3609 represents an intronic
FI exon 3791..3923
FI /tag= q
FI /number= 6
FI intron 3924..4439
FI /tag= r
FI /number= 6
FI misc_difference 4185
FI /tag= s
FI /note= "n at position 4185 represents an intronic
FI exon 4440..4635
FI /tag= t
FI /number= 7
FI intron 4636..5143
FI /tag= u
FI /number= 7
FI misc_difference 4857
FI /tag= v
FI /note= "n at position 4857 represents an intronic
FI exon 5144..5262
FI /tag= w
FI /number= 8
FI intron 5263..5475
FI /tag= x
FI /number= 8
FI misc_difference 5369
FI /tag= y
FI /note= "n at position 5369 represents an intronic
FI exon 5476..6441
FI /tag= z

```

```

FT XX                               /number= 9
PN XX
PD XX WO9811254-A1.
PE XX 19-MAR-1998.
PR XX 10-SEP-1997; 97WO-US16037.
PR XX 30-OCT-1996; 96US-0029679.
PR XX 10-SEP-1996; 96US-0025719.
PR XX 02-OCT-1996; 96US-0028056.
PA XX (ARCH-) ARCH DEV CORP.
XX
PI Bell GT, Furuta H, Horikawa Y, Kaisaki PJ, Menzel S;
PI Oda N, Yamagata K;
XX
DR WPI; 1998-271667/24.
DR P-PSDB; AAW71581 and AAW71586.
XX
PI Isolated nucleic acid encoding hepatocyte nuclear factor 1-alpha and
PI 1-beta - useful for detecting susceptibility for non-insulin
PT dependent diabetes, especially maturity-onset diabetes of the young
PS Claim 89; Fig 27A-I; 363pp; English.
XX
CC This is a partial nucleotide sequence of the gene (TCF2) encoding
CC human hepatocyte nuclear factor 1 beta (HNF-1 beta, see AAW71581), a
CC homeodomain-containing transcription factor. Mutations in this
CC gene are indicative of a propensity to diabetes mellitus. The
CC invention concerns the identification of genes responsible for
CC non-insulin dependent diabetes mellitus (NIDDM) for use in
CC diagnostics and therapeutics. It demonstrates that the MODY3 locus
CC is the HNF-1 alpha gene (see AAV52625). It also demonstrates that the
CC beta gene (see also AAV52730) and the MODY1 locus is the HNF-4 alpha
CC gene (see AAV52687). Analysis of mutations in these HNF genes can be
CC diagnostic for diabetes. The invention also contemplates methods of
CC screening for modulators of HNF function utilising HNF nucleic acids
CC or polypeptides, the modulators being useful for treating diabetes
CC by modulating HNF function in an animal.
XX
SQ Sequence 6441 BP; 1572 A; 1827 C; 1530 G; 1504 T; 8 other;

Query Match      2.8%; Score 17; DB 19; Length 6441;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 245 ccgcaggaggaggagg 261
   |||||
Db 2685 ccgcaggaggaggagg 2705

RESULT 43
AAV52207
ID AAV52207 standard; DNA; 16535 BP.
XX
AC AAV52207;
XX
DT 23-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae genome fragment SEQ ID NO:74.
XX
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
OS Streptococcus pneumoniae.
XX
PN WO9818931-A2.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-US19588.

```

```

XX
PR 31-OCT-1996; 96US-0029560.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
PI Kunsch CA, Rasmussen CA;
XX
DR WPI; 1998-272225/24.
XX
CC Computer-readable medium with recorded Streptococcus pneumoniae
PI polynucleotide sequences - useful in diagnostic kits and assays, and
PI pharmaceutical compositions and vaccines for Streptococcus
PI pneumoniae
XX
PS Claim 1; Page 608-617; 1409pp; English.
XX
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC recorded on it, or a representative fragment or a sequence at least 95%
CC identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in
CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC Streptococcus pneumoniae. The present invention also describes an
CC isolated nucleic acid molecule encoding a homologue of any of the
CC fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the
CC nucleic acid molecule is produced by a process comprising: (a) screening
CC a genomic DNA library using as a probe a target sequence defined by any
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC library which contain sequences that hybridise to the target sequence and
CC isolating the nucleic acid molecules from the members; or (b) isolating
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC molecules whose nucleotide sequence is homologous to amplification
CC primers derived from the fragment of the S. pneumoniae genome to prime
CC the amplification and isolating the amplified sequences. The computer
CC readable medium can be used in a computer-based system for identifying
CC fragments of the S. pneumoniae genome of commercial importance, or
CC expression modulating fragments of the S. pneumoniae genome. Products
CC from the present invention can be used in diagnosis kits and assays, and
CC pharmaceutical compositions and vaccines for S. pneumoniae.
XX
SQ Sequence 16535 BP; 5001 A; 3914 C; 3102 G; 4518 T; 0 other;

```

```

Query Match      2.8%; Score 17; DB 19; Length 16535;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 474 ttcaaggatgccaaaga 490
   |||||
Db 451 ttcaaggatgccaaaga 467

RESULT 44
AAF88315
ID AAF88315 standard; DNA; 45624 BP.
XX
AC AAF88315;
XX
DT 28-AUG-2001 (first entry)
XX
DE S. spinosa DNA fragment SEQ ID 4.
XX
KW Forosamin; trimethylthamose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; ds.
XX
OS Saccharopolyspora spinosa.
XX
PN DE19557268-A1.
XX
PD 08-MAR-2001.
XX
PF 29-NOV-1999; 99DE-1057268.

```

XX 27-AUG-1999; 99DE-1040596.
 PR (FARB) BAYER AG.
 PA Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
 XX WPI; 2001-267102/28.
 XX
 XX New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
 PT recombinant production of insecticidal spinosyns and their derivatives
 PT
 XX
 XX Claim 7: Page 58-74; 354pp; German.
 XX
 XX This invention describes a novel method nucleic acid (I) and its encoded
 CC polypeptide (II) containing at least one region that encodes an enzymatic
 CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
 CC identify, inactivate or modulate genes involved in the biosynthesis of
 CC (II); (ii) to generate a library of polyketide synthases; (iii) for
 CC adding forosamine or trimethylrhannose to a spinosyn or polyketide
 CC aglycone; and (iv) for recombinant production of the corresponding
 CC enzymes, which are used for production of (II), their precursors or
 CC derivatives, including production of transgenic plants that express (II)
 CC and thus have increased resistance to insects. (I) are also useful as
 CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
 CC macrolides with insecticidal, but not antibacterial, activity, and can
 CC also be used to raise specific antibodies, useful for identifying
 CC expression clones in a gene bank. Cells transformed with (I) may produce
 CC (II) at significantly increased levels or produce new derivatives of
 CC (II). This sequence represents a genomic DNA fragment of the S. spinosa
 CC genome which contains the coding regions for proteins involved in
 CC forosamine and trimethylrhannose biosynthesis.
 XX
 SQ Sequence 45624 BP; 7933 A; 14913 C; 14940 G; 7838 T; 0 other;

Query Match 2.8%; Score 17; DB 22; Length 45624;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 329 gcgcacctcctggagatcg 345
 |||||
 Db 6789 gcgcacctcctggagatcg 6805

RESULT 45
 AAF88312/C
 ID AAF88312 standard; DNA; 50000 BP.
 XX
 AC AAF88312;

DT 28-AUG-2001 (first entry)
 XX
 DE S. spinosa DNA fragment SEQ ID 1.
 XX

KW Forosamine; trimethylrhannose; polyketide synthase; biosynthesis;
 KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
 KW macrolide; insecticidal; ds.
 XX

OS Saccharopolyspora spinosa.
 XX

PN DE19957268-A1.
 XX

PD 08-MAR-2001.
 XX

PF 29-NOV-1999; 99DE-1057268.
 XX

PR 27-AUG-1999; 99DE-1040596.
 XX

PA (FARB) BAYER AG.
 XX

PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;

XX
 DR

WPI; 2001-267102/28.

XX

New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
 recombinant production of insecticidal spinosyns and their derivatives

PT

XX

Claim 7: Page 14-31; 354pp; German.

XX

This invention describes a novel method nucleic acid (I) and its encoded
 polypeptide (II) containing at least one region that encodes an enzymatic
 activity involved in biosynthesis of spinosyns. (I) are used (i) to
 identify, inactivate or modulate genes involved in the biosynthesis of
 (II); (ii) to generate a library of polyketide synthases; (iii) for
 adding forosamine or trimethylrhannose to a spinosyn or polyketide
 aglycone; and (iv) for recombinant production of the corresponding
 enzymes, which are used for production of (II), their precursors or
 derivatives, including production of transgenic plants that express (II)
 and thus have increased resistance to insects. (I) are also useful as
 markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
 macrolides with insecticidal, but not antibacterial, activity, and can
 also be used to raise specific antibodies, useful for identifying
 expression clones in a gene bank. Cells transformed with (I) may produce
 (II) at significantly increased levels or produce new derivatives of
 (II). This sequence represents a genomic DNA fragment of the S. spinosa
 genome which contains the coding regions for proteins involved in
 forosamine, trimethylrhannose and polyketide synthase biosynthesis.

SQ Sequence 50000 BP; 8484 A; 16290 C; 16535 G; 8691 T; 0 other;

Query Match 2.8%; Score 17; DB 22; Length 50000;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 329 gcgcacctcctggagatcg 345
 |||||

Db 38836 GCGCATCTGAGATCG 38820

Search completed: April 30, 2002, 10:51:14
 Job time: 10743 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2002, 19:49:59 ; Search time 888.17 Seconds
(without alignments)
927.884 Million cell updates/sec

Title: US-09-248-178-56

Perfect score: 480

Sequence: 1 ggctttgaagcattttgtc.....gcaacaaaaaaaaaaaaa 480

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

1:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	480	100.0	480	20	AAx84202
2	480	100.0	480	21	AAx79431
3	480	100.0	489	20	AAx00704
4	480	100.0	489	20	AAx00644
5	480	100.0	578	21	AAx79470
6	480	100.0	578	22	AAx16724
7	478.4	99.7	553	19	AAx31993
8	478.4	99.7	553	20	AAx55582
9	476.4	99.2	497	20	AAx40783

10	174.8	58.7	544	21	AAx91765	Human breast speci
11	469.4	57.8	471	21	AAx65099	Membrane-bound pro
12	469.4	57.8	471	22	AAx44245	Human PRO1160 (UNQ
13	469.4	57.8	471	22	AAx97492	Human angiogenesis
14	468.6	57.6	488	22	AAx44870	Human breast cance
15	465.2	56.9	482	19	AAx31992	BS106 polynucleoti
16	465.2	56.9	482	20	AAx55581	BS106 consensus po
17	464.4	56.8	472	22	AAx18673	Human BS106 gene-s
18	464.4	56.8	472	22	AAx18674	Human BS106 gene-s
19	450	53.8	471	21	AAx54120	Breast cancer prot
20	450	53.8	471	22	AAx85500	Nucleotide sequenc
21	307	64.0	501	23	AAx87721	BS106 polynucleoti
22	304.4	63.4	308	19	AAx31990	BS106 polynucleoti
23	304.4	63.4	308	20	AAx55579	Human BS106 gene-s
24	304.4	63.4	308	22	AAx18670	Human BS106 gene-s
25	288.4	60.1	292	22	AAx18671	Breast cancer prot
26	273	56.9	273	21	AAx54121	cDNA encoding SRT
27	259.2	54.0	264	22	AAx93398	cDNA encoding SRT
28	259.2	54.0	266	22	AAx93397	BS106 polynucleoti
29	250.8	52.2	291	22	AAx93397	BS106 clone 166288
30	206.4	43.0	229	19	AAx31989	BS106 clone 120981
31	206.4	43.0	229	20	AAx55578	Human BS106 gene-s
32	196	40.8	197	19	AAx31991	Human BS106 gene-s
33	196	40.8	197	20	AAx55580	Human BS106 gene-s
34	196	40.8	197	22	AAx18672	Human BS106 gene-s
35	188.4	35.2	201	22	AAx18669	Human BS106 gene-s
36	96	20.0	660	23	AAx87720	DNA encoding novel
37	70.2	14.6	397	20	AAx89891	Spinocerebellar at
38	65.4	13.6	477	21	AAx44307	Human SCA7 genomic
39	65	13.5	234	16	AAx84832	Spinocerebellar at
40	64.8	13.5	1326	23	AAx83104	DNA encoding novel
41	64.8	13.5	3376	16	AAx75166	AF-9 cDNA. Homo s
42	64.6	13.5	203	19	AAx30271	Glutamine rich reg
43	64.6	13.5	203	19	AAx17226	SCA2 gene CAG repe
44	64.6	13.5	1037	21	AAx59242	Exons E, C and A o
45	64.6	13.5	1472	21	AAx59241	Exons D, C, B and

ALIGNMENTS

RESULT 1
AAx84202
ID AAx84202 standard; cDNA; 480 BP.
XX AAx84202:
XX
XX 08-SEP-1999 (first entry)
XX
XX DNA encoding human breast tumour protein immunogenic fragment.
XX
XX Breast tumour protein; immunogenic fragment; vaccine; detection;
XX Breast cancer development; therapy; ss.
XX
XX Homo sapiens.
XX
XX W09933869-A2.
XX
XX 08-JUL-1999.
XX
XX 22-DEC-1998; 98WO-US27416.
XX
XX 17-JUL-1998; 98US-0118627.
XX 24-DEC-1997; 97US-0598253.
XX 24-DEC-1997; 97US-0598255.
XX 17-JUL-1998; 98US-0118554.
XX (CORI-) CORIXA CORP.
XX
XX Reed SG, Xu J.
XX WPI; 1999-405486/34.
XX

PT New breast tumour protein genes used, in vaccines for immunotherapy;
 PT or for diagnosis of breast cancer
 XX
 PS
 PS
 XX
 XX
 XX
 CC This sequence encodes a human breast tumour protein immunogenic fragment
 CC of the invention. The polypeptides or nucleic acids encoding them are
 CC useful in vaccines and pharmaceutical compositions for manufacture of
 CC medicaments for inhibiting the development of breast cancer in a patient.
 CC They can also be used to treat breast cancer. Antibodies against these
 CC polypeptides can be used to detect and monitor progression of breast
 CC cancer in patients. Primers and probes derived from the polynucleotides
 CC encoding the breast proteins are useful for detection of breast cancer.
 CC Peripheral blood cells from a patient incubated in the presence of at
 CC least one polypeptide, such that T cells proliferate, are useful in
 CC manufacture of a medicament for treating breast cancer in a patient.
 CC Antigen presenting cells incubated in the presence of at least one
 CC polypeptide are also useful for treating breast cancer.
 XX
 SQ Sequence 480 BP; 121 A; 132 C; 87 G; 140 T; 0 other;

Query Match 100.0%; Score 480; DB 20; Length 480;
 Best Local Similarity 100.0%; Pred. No. 1.9e-109;
 Matches 480; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ggccttgaagcattttgtctgctccctgagcttcaggtccaccaccatgaagtcttta 60
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1 ggccttgaagcattttgtctgctccctgagcttcaggtccaccaccatgaagtcttta 60
 QY 61 gcagtcctgtaactctggagggtttccattcttctgtctctgtccagagaatccgacaaca 120
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 61 gcagtcctgtaactctggagggtttccattcttctgtctctgtccagagaatccgacaaca 120
 QY 121 gctgtccagctgacagtgatccagctgagctgctctgctgagtgatgaagccctgatgct 180
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 121 gctgtccagctgacagtgatccagctgagctgctctgctgagtgatgaagccctgatgct 180
 QY 181 gaaccactgtctgcaaccactgcgaccactgctgctctaccactgcaaccaccgct 240
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 181 gaaccactgtctgcaaccactgcgaccactgctgctctaccactgcaaccaccgct 240
 QY 241 gcttccaccctgctgtaagacattccagttttaccaccaaatgggttggggatctcccg 300
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 241 gcttccaccctgctgtaagacattccagttttaccaccaaatgggttggggatctcccg 300
 QY 301 aatggcagagtgtccctgagatgggaatcagcttgagctgctctgcaattggtcacaaact 360
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 301 aatggcagagtgtccctgagatgggaatcagcttgagctgctctgcaattggtcacaaact 360
 QY 361 attcatgcttctgtgatttcacactgcttaccctgctacgatatccccctttatct 420
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 361 attcatgcttctgtgatttcacactgcttaccctgctacgatatccccctttatct 420
 QY 421 ctaatcagttatttttttccaaataaaaaataactatgagcaacaaaaaataaaaaaaa 480
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 421 ctaatcagttatttttttccaaataaaaaataactatgagcaacaaaaaataaaaaaaa 480

RESULT 2
 AAC79431
 ID AAC79431 standard; cDNA; 480 BP.
 XX AC AAC79431;
 XX AC AAC79431;
 DT 07-FEB-2001 (first entry)
 XX cDNA sequence of human breast tumour clone 1016F8.
 DE Human; breast tumour antigen; cytostatic; immunotherapy;
 XX breast cancer; vaccine; ss.
 KW Homo sapiens.
 XX
 OS

XX WO2000061756-A2.
 PH 19-OCT-2000.
 XX
 XX 10-APR-2000; 20JUN00-US09688.
 XX
 XX 05-APR-1999; JUS-0285950.
 PR 02-JUL-1999; JUS-0346327.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 XX Reed SG, Xu J, Dillon DC;
 PI WPI; 2000-638563/61.
 DR
 XX A novel isolated polypeptide comprising an immunogenic portion of a
 PT breast cancer protein useful in the detection and treatment of breast
 PT cancer -
 PT
 XX
 PS Claim 4; Page 74; 95pp; English.
 XX
 CC The present sequence was isolated from a breast tumour cDNA library. It
 CC is provided in a specification relating to compounds for immunotherapy
 CC and diagnosis of breast cancer. Breast tumour antigens and the
 CC polynucleotides that encode them may be used in the production of a
 CC pharmaceutical composition to be used in the treatment of breast cancer.
 CC Proliferated T cells and incubated antigen presenting cells are also
 CC required. The polypeptides and polynucleotides may also be used to
 CC produce a vaccine.
 XX
 SQ Sequence 480 BP; 121 A; 132 C; 87 G; 140 T; 0 other;

Query Match 100.0%; Score 480; DB 21; Length 480;
 Best Local Similarity 100.0%; Pred. No. 1.9e-109;
 Matches 480; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ggccttgaagcattttgtctgctccctgagcttcaggtccaccaccatgaagtcttta 60
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1 ggccttgaagcattttgtctgctccctgagcttcaggtccaccaccatgaagtcttta 60
 QY 61 gcagtcctgtaactctggagggtttccattcttctgtctctgtccagagaatccgacaaca 120
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 61 gcagtcctgtaactctggagggtttccattcttctgtctctgtccagagaatccgacaaca 120
 QY 121 gctgtccagctgacagtgatccagctgagctgctctgctgagtgatgaagccctgatgct 180
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 121 gctgtccagctgacagtgatccagctgagctgctctgctgagtgatgaagccctgatgct 180
 QY 181 gaaccactgtctgcaaccactgcgaccactgctgctctaccactgcaaccaccgct 240
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 181 gaaccactgtctgcaaccactgcgaccactgctgctctaccactgcaaccaccgct 240
 QY 241 gcttccaccctgctgtaagacattccagttttaccaccaaatgggttggggatctcccg 300
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 241 gcttccaccctgctgtaagacattccagttttaccaccaaatgggttggggatctcccg 300
 QY 301 aatggcagagtgtccctgagatgggaatcagcttgagctgctctgcaattggtcacaaact 360
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 301 aatggcagagtgtccctgagatgggaatcagcttgagctgctctgcaattggtcacaaact 360
 QY 361 attcatgcttctgtgatttcacactgcttaccctgctacgatatccccctttatct 420
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 361 attcatgcttctgtgatttcacactgcttaccctgctacgatatccccctttatct 420
 QY 421 ctaatcagttatttttttccaaataaaaaataactatgagcaacaaaaaataaaaaaaa 480
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 421 ctaatcagttatttttttccaaataaaaaataactatgagcaacaaaaaataaaaaaaa 480
 RESULT 3
 AAX00704

ID AAX00704 standard; DNA; 489 BP.
 XX AAX00704;
 AC AAX00704;
 XX AAX00704;
 DT 25-MAR-1999 (first entry)
 XX Human secreted protein gene 34 clone HMQAJ64.
 DE Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; gastritis; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX Homo sapiens.
 OS Homo sapiens.
 PN WO9842738-A1.
 XX 01-OCT-1998.
 PD 01-OCT-1998.
 XX 19-MAR-1998; 98WO-US05311.
 XX 30-MAY-1997; 97US-0050937.
 PR 21-MAR-1997; 97US-0041276.
 PR 21-MAR-1997; 97US-0041277.
 PR 21-MAR-1997; 97US-0041281.
 PR 21-MAR-1997; 97US-0042344.
 PR 30-MAY-1997; 97US-0048069.
 PR 30-MAY-1997; 97US-0048094.
 PR 30-MAY-1997; 97US-0048095.
 PR 30-MAY-1997; 97US-0048096.
 PR 30-MAY-1997; 97US-0048099.
 PR 30-MAY-1997; 97US-0048131.
 PR 30-MAY-1997; 97US-0048135.
 PR 30-MAY-1997; 97US-0048154.
 PR 30-MAY-1997; 97US-0048160.
 PR 30-MAY-1997; 97US-0048186.
 PR 30-MAY-1997; 97US-0048187.
 PR 30-MAY-1997; 97US-0048188.
 PR 30-MAY-1997; 97US-0048350.
 PR 30-MAY-1997; 97US-0048351.
 PR 30-MAY-1997; 97US-0048352.
 PR 30-MAY-1997; 97US-0048355.
 PR 05-AUG-1997; 97US-0054804.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Brewer LA, Duan R, Ebner R, Fertie AM, Florence KA;
 PI Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS;
 PI Rosen CA, Ruben SM, Shi Y, Young P;
 XX WPI; 1999-070066/06.
 DR P-PSDB; AAW67900.
 XX New isolated human genes and the secreted polypeptides they encode
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX Claim 1; Page 254; 385pp; English.
 XX This sequence represents a nucleic acid molecule which encodes a
 CC secreted human protein. The gene number, and the clone it is derived
 CC from, are detailed in the descriptor line. The gene can be used to
 CC generate fusion proteins by linking to the gene to a human immunoglobulin
 CC Fc portion (e.g. AAX00602) for increasing the stability of the fused
 CC protein as compared to the human protein only.
 CC The invention relates to 87 novel genes and their fragments (nucleic
 CC acid sequences: AAX00611-X00724; amino acid sequences AAW67807-W68004)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological

CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 87
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAX00611 for described uses).
 XX Sequence 489 BP; 124 A; 134 C; 91 G; 140 T; 0 other;
 SQ query Match 100.0%; Score 480; DB 20; Length 489;
 Best Local Similarity 100.0%; Pred. No. 1.5e-109;
 Matches 480; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ggtttgaaacattttgtctgctccctgattctcagtcacccaccatgaagttcttta 60
 DE 10 ggtttgaaacattttgtctgctccctgattctcagtcacccaccatgaagttcttta 69
 QY 61 gcagtcctgactcttgggagtttccatcttctgtctctgctccagaaatccgacaaca 120
 DE 70 gcagtcctgactcttgggagtttccatcttctgtctctgctccagaaatccgacaaca 129
 QY 121 gctgctccagctgacacgtatccagctactggtctctgctgctgattgaagccctgatgct 180
 DE 130 gctgctccagctgacacgtatccagctactggtctctgctgctgattgaagccctgatgct 189
 QY 181 gaaacacactgctgcaacacactgcacacactgctgctctcctaccactgcaacacacgct 240
 DE 190 gaaacacactgctgcaacacactgcacacactgctgctctcctaccactgcaacacacgct 249
 QY 241 gcttctaccactgctgtaagacattccagttttaccacaaatgggttggggatctcccg 300
 DE 250 gcttctaccactgctgtaagacattccagttttaccacaaatgggttggggatctcccg 309
 QY 301 aatggtagatgctgctccctgagatggaatcagcttctgctgcaatgggtcacaact 360
 DE 310 aatggtagatgctgctccctgagatggaatcagcttctgctgcaatgggtcacaact 369
 QY 361 attcatgcttctgctgatttcacacactacttaccctgctcctacgatatccctttatct 420
 DE 370 attcatgcttctgctgatttcacacactacttaccctgctcctacgatatccctttatct 429
 QY 421 ctatcagtttatttcttcttcaaaaaataaactatgagcaacaaaaaataaaaaa 480
 DE 430 ctatcagtttatttcttcttcaaaaaataaactatgagcaacaaaaaataaaaaa 489
 RE-SULT 4
 AAX00644
 ID AAX00644 standard; DNA; 489 BP.
 XX AAX00644;
 AC AAX00644;
 XX AAX00644;
 DT 25-MAR-1999 (first entry)
 XX Human secreted protein gene 34 clone HMQAJ64.
 DE Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; gastritis; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX Homo sapiens.
 OS Homo sapiens.
 PN WO9842738-A1.
 XX 01-OCT-1998.
 PD 01-OCT-1998.
 XX 19-MAR-1998; 98WO-US05311.
 XX 30-MAY-1997; 97US-0050937.
 PR 21-MAR-1997; 97US-0041276.
 PR 21-MAR-1997; 97US-0041277.
 PR 21-MAR-1997; 97US-0041281.
 PR 21-MAR-1997; 97US-0042344.
 PR 30-MAY-1997; 97US-0048069.
 PR 30-MAY-1997; 97US-0048094.
 PR 30-MAY-1997; 97US-0048095.
 PR 30-MAY-1997; 97US-0048096.
 PR 30-MAY-1997; 97US-0048099.
 PR 30-MAY-1997; 97US-0048131.
 PR 30-MAY-1997; 97US-0048135.
 PR 30-MAY-1997; 97US-0048154.
 PR 30-MAY-1997; 97US-0048160.
 PR 30-MAY-1997; 97US-0048186.
 PR 30-MAY-1997; 97US-0048187.
 PR 30-MAY-1997; 97US-0048188.
 PR 30-MAY-1997; 97US-0048350.
 PR 30-MAY-1997; 97US-0048351.
 PR 30-MAY-1997; 97US-0048352.
 PR 30-MAY-1997; 97US-0048355.
 PR 05-AUG-1997; 97US-0054804.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Brewer LA, Duan R, Ebner R, Fertie AM, Florence KA;
 PI Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS;
 PI Rosen CA, Ruben SM, Shi Y, Young P;
 XX WPI; 1999-070066/06.
 DR P-PSDB; AAW67900.
 XX New isolated human genes and the secreted polypeptides they encode
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX Claim 1; Page 254; 385pp; English.
 XX This sequence represents a nucleic acid molecule which encodes a
 CC secreted human protein. The gene number, and the clone it is derived
 CC from, are detailed in the descriptor line. The gene can be used to
 CC generate fusion proteins by linking to the gene to a human immunoglobulin
 CC Fc portion (e.g. AAX00602) for increasing the stability of the fused
 CC protein as compared to the human protein only.
 CC The invention relates to 87 novel genes and their fragments (nucleic
 CC acid sequences: AAX00611-X00724; amino acid sequences AAW67807-W68004)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological

30-MAY-1997; 97US-0050937.
21-MAR-1997; 97US-0041276.
21-MAR-1997; 97US-0041277.
21-MAR-1997; 97US-0041281.
21-MAR-1997; 97US-0042344.
30-MAY-1997; 97US-0048069.
30-MAY-1997; 97US-0048094.
30-MAY-1997; 97US-0048095.
30-MAY-1997; 97US-0048096.
30-MAY-1997; 97US-0048099.
30-MAY-1997; 97US-0048131.
30-MAY-1997; 97US-0048135.
30-MAY-1997; 97US-0048154.
30-MAY-1997; 97US-0048160.
30-MAY-1997; 97US-0048186.
30-MAY-1997; 97US-0048187.
30-MAY-1997; 97US-0048188.
30-MAY-1997; 97US-0048350.
30-MAY-1997; 97US-0048351.
30-MAY-1997; 97US-0048352.
30-MAY-1997; 97US-0048355.
05-AUG-1997; 97US-0054804.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Brewer LA, Duan R, Ebner R, Ferrie AM, Florence KA;
PI Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS;
PI Rosen CA, Ruben SM, Shi Y, Young P;
XX
XX WPI; 1999-070066/06.
DR P-PSDB; AAW67840.
XX
XX New isolated human genes and the secreted polypeptides they encode -
PI useful for diagnosis and treatment of e.g. cancers, neurological
PI disorders, immune diseases, inflammation or blood disorders
XX
XX Claim 1; Page 198; 385pp; English.
XX
XX This sequence represents a nucleic acid molecule which encodes a
CC secreted human protein. The gene number, and the clone it is derived
CC from, are detailed in the descriptor line. The gene can be used to
CC generate fusion proteins by linking to the gene to a human immunoglobulin
CC Fc portion (e.g. AAX00602) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 87 novel genes and their fragments (nucleic
CC acid sequences: AAX00611-X00724; amino acid sequences AAW67807-W68004)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 87
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAX00611 for described uses).
XX
XX Sequence 489 BP; 125 A; 134 C; 87 G; 143 T; 0 other;

Query Match 100.0%; Score 480; DB 20; Length 489;
Best Local Similarity 100.0%; Pred. No. 1.9e-109;
Matches 480; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggtcttgaagcattttgtctgtccctgattcttcagggtccaccacgaagtcttta 60
DB 7 ggtcttgaagcattttgtctgtccctgattcttcagggtccaccacgaagtcttta 66
QY 61 gaagtcctggtactcttgggagttccatcttctgtctgtccagaaatccgacaca 120
DB 67 gaagtcctggtactcttgggagttccatcttctgtctgtccagaaatccgacaca 126
QY 121 gctgtccagctgacacgtatccagactactggtctgtctgtgatgaagccctgatgct 180
DB 127 gctgtccagctgacacgtatccagactactggtctgtctgtgatgaagccctgatgct 186

QY 181 gaaccactgctgtgcacacactgcgaccactgctgtctctaccactgcacacccgct 240
DB 187 gaaccactgctgtgcacacactgcgaccactgctgtctctaccactgcacacccgct 246
QY 241 gcttctaccactgctcgtaaagacattccagttttaccacaaatgggttggggatctcccg 300
DB 247 gcttctaccactgctcgtaaagacattccagttttaccacaaatgggttggggatctcccg 306
QY 301 aatggttagtgatgtgctccctgagatggaatcagcttgagttcttgcatttggatcacaact 360
DB 307 aatggttagtgatgtgctccctgagatggaatcagcttgagttcttgcatttggatcacaact 366
QY 361 attcargcttctgtgatttccatcactacttgccttgcctacatcccttttact 420
DB 367 attcargcttctgtgatttccatcactacttgccttgcctacatcccttttact 426
QY 421 ctaatcagtttattttcttccaaataaaaaataaactatgagcaacaaaaa 480
DB 427 ctaatcagtttattttcttccaaataaaaaataaactatgagcaacaaaaa 486
RESULT 5
AAC79470
ID AAC79470 standard; cDNA; 578 BP.
XX
XX AC AAC79470;
XX
XX AC AAC79470;
DT 07-FEB-2001 (first entry)
XX
XX cDNA sequence of human breast tumour clone B511S.
DE Human; breast tumour antigen; cytostatic; immunotherapy;
XX Human; breast tumour antigen; cytostatic; immunotherapy;
KW breast cancer; vaccine; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200061756-A2.
PN
XX
XX 19-OCT-2000.
PD
XX
XX 10-APR-2000; 2000WO-US09688.
PF
XX
XX 09-APR-1999; 99US-0288950.
PR
XX
XX 02-JUL-1999; 99US-0346327.
PR
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Reed SG, Xu J, Dillon DC;
PI
XX
XX WPI; 2000-638568/61.
DR
XX
XX P-PSDB; AAS28525.
XX
XX A novel isolated polypeptide comprising an immunogenic portion of a
PI breast cancer protein useful in the detection and treatment of breast
PI cancer -
PI
XX
XX Claim 4; Page 88; 95pp; English.
PS
XX
XX The present sequence was isolated from a breast tumour cDNA library. It
CC is provided in a specification relating to compounds for immunotherapy
CC and diagnosis of breast cancer. Breast tumour antigens and the
CC polynucleotides that encode them may be used in the production of a
CC pharmaceutical composition to be used in the treatment of breast cancer.
CC Proliferated I cells and incubated antigen presenting cells are also
CC required. The polypeptides and polynucleotides may also be used to
CC produce a vaccine.
CC
XX
XX Sequence 578 BP; 206 A; 137 C; 88 G; 147 T; 0 other;

Query Match 100.0%; Score 480; DB 21; Length 578;
Best Local Similarity 100.0%; Pred. No. 2e-109;
Matches 480; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ggctttgaagcatttttctgtctgtccctgatctcttcaggtcaccaccatgaagttctta	60
Db	15	ggctttgaagcatttttctgtctgtccctgatctcttcaggtcaccaccatgaagttctta	74
QY	61	gcagtcctggtactcttgaggagttcccatctttcttggtctctgcccagaaatccgacaca	120
Db	75	gcagtcctggtactcttgaggagttcccatctttcttggtctctgcccagaaatccgacaca	134
QY	121	gctgtctcagctgacacgtatccagctactggtctctgtgatgatgaagccctctgatgt	180
Db	135	gctgtctcagctgacacgtatccagctactggtctctgtgatgatgaagccctctgatgt	194
QY	181	gaacacactgctctgaacactgcgacacactgctctcttaccactgaacacacccgt	240
Db	195	gaacacactgctctgaacactgcgacacactgctctcttaccactgaacacacccgt	254
QY	241	gctctaccactgctctgaagacattccagttttaccacaaatgggttggtatctcccg	300
Db	255	gctctaccactgctctgaagacattccagttttaccacaaatgggttggtatctcccg	314
QY	301	aatggtagagtggtccctcgagatgggaatcagcttgagctcttcgcaatgggtcacaact	360
Db	315	aatggtagagtggtccctcgagatgggaatcagcttgagctcttcgcaatgggtcacaact	374
QY	361	attcatgctctctgtagttcttcacaaactacttacccttgcctacgataccccctttatct	420
Db	375	attcatgctctctgtagttcttcacaaactacttacccttgcctacgataccccctttatct	434
QY	421	ctaatacagttattttcttttccaaataaaaaataactatgatgacacacacacacacacac	480
Db	435	ctaatacagttattttcttttccaaataaaaaataactatgatgacacacacacacacacac	494
RESULT 6			
ID	AAI67224 standard; cDNA; 578 BP.		
XX			
AC	AAI67224;		
DT	11-FEB-2002 (first entry)		
XX			
DE	B511S cDNA sequence.		
XX			
KW	Genetic subtraction; DNA microarray analysis; polymerase chain reaction;		
KW	cancer; B511S; ss.		
XX			
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	63..335	
FT	/*tag= a		
XX			
PN	WO200175171-A2.		
XX			
PD	11-OCT-2001.		
XX			
PF	02-APR-2001; 2001WO-US10631.		
XX			
PR	03-APR-2000; 2000US-194241P.		
PR	20-JUL-2000; 2000US-219862P.		
PR	27-JUL-2000; 2000US-221300P.		
PR	18-DEC-2000; 2000US-256592P.		
XX			
PA	(CORI-) CORIXA CORP.		
XX			
PI	Houghton RL, Dillon DC, Molesh DA, Xu J, Zehentner B, Persing DH;		
XX			
DR	WPI; 2001-626449/72.		
DR	P-PSDB; AAG65988.		
XX			
PT	Identifying tissue (tumour)-specific polynucleotides overexpressed in		
PT	tissue of interest as compared to control tissue, for detecting cancer		

PI	cells in patient, comprises DNA microarray analysis or quantitative		
PI	polymerase chain reaction -		
XX			
PS	Claim 4; Page 116; 127pp: English.		
XX			
CC	The invention relates to identifying tissue-specific polynucleotides (P)		
CC	that involves performing a genetic subtraction to identify pool of (P)		
CC	from tissue of interest (TI), performing DNA microarray analysis to		
CC	identify first subset of polynucleotides (SPI) at least 2-fold over		
CC	expressed in TI, and performing quantitative polymerase chain reaction		
CC	(PCR) analysis on SPI to identify second subset of (P). The method is		
CC	useful for determining the presence or absence of a cancer cell in a		
CC	patient, monitoring the progression of cancer in a patient using a		
CC	biological sample such as blood, serum, lymph nodes, bone marrow, sputum,		
CC	urine or a tumour biopsy sample. The methods are useful for determining		
CC	the presence or absence of or monitoring progression of prostate, breast,		
CC	colon, ovarian, lung, head and neck, lymphoma, leukemia, melanoma, liver,		
CC	gastric, kidney, bladder, pancreatic or endometrial cancer. The present		
CC	sequence represents B511S cDNA.		
XX			
SQ	Sequence 578 BP; 206 A; 137 C; 88 G; 147 T; 0 other;		
Query Match 100.0%; Score 480; DB 22; Length 578;			
Best Local Similarity 100.0%; Pred. No. 2e-109;			
Matches 480; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	ggctttgaagcatttttctgtctgtccctgatctcttcaggtcaccaccatgaagttctta	60
Db	15	ggctttgaagcatttttctgtctgtccctgatctcttcaggtcaccaccatgaagttctta	74
QY	61	gcagtcctggtactcttgaggagttcccatctttcttggtctctgcccagaaatccgacaca	120
Db	75	gcagtcctggtactcttgaggagttcccatctttcttggtctctgcccagaaatccgacaca	134
QY	121	gctgtctcagctgacacgtatccagctactggtctctgtgatgatgaagccctctgatgt	180
Db	135	gctgtctcagctgacacgtatccagctactggtctctgtgatgatgaagccctctgatgt	194
QY	181	gaacacactgctctgaacactgcgacacactgctctcttaccactgaacacacccgt	240
Db	195	gaacacactgctctgaacactgcgacacactgctctcttaccactgaacacacccgt	254
QY	241	gctctaccactgctctgaagacattccagttttaccacaaatgggttggtatctcccg	300
Db	255	gctctaccactgctctgaagacattccagttttaccacaaatgggttggtatctcccg	314
QY	301	aatggtagagtggtccctcgagatgggaatcagcttgagctcttcgcaatgggtcacaact	360
Db	315	aatggtagagtggtccctcgagatgggaatcagcttgagctcttcgcaatgggtcacaact	374
QY	361	attcatgctctctgtagttcttcacaaactacttacccttgcctacgataccccctttatct	420
Db	375	attcatgctctctgtagttcttcacaaactacttacccttgcctacgataccccctttatct	434
QY	421	ctaatacagttattttcttttccaaataaaaaataactatgatgacacacacacacacac	480
Db	435	ctaatacagttattttcttttccaaataaaaaataactatgatgacacacacacacacac	494
RESULT 7			
AAV31993			
ID	AAV31993 standard; DNA; 553 BP.		
XX			
AC	AAV31993;		
DT	25-SEP-1998 (first entry)		
XX			
DE	BS106 polynucleotide clone 1662885 (ii).		
XX			
KW	ss; human; BS106; breast tissue gene; breast cancer; detection marker.		
XX			
OS	Homo sapiens.		

[illegible]

```
Db 390 attcagctccggtgattctcacaactcacttactctgctcctacgataccccctttatct 449
Qy 421 ctaatcagttattttcttctcaataaaaaataaactatgagcaacaaaaaaatcttctta 480
Db 450 ctaatcagttattttcttctcaataaaaaataaactatgagcaacaaaaaaatcttctta 509

RESULT 9
ID AAZ40783
XX AAZ40783 standard; DNA; 497 BP.
AC AAZ40783;
XX
DT 18-JAN-2000 (first entry)
XX
DE Secreted protein EST coding sequence 108-002-5-0-F3-FL.
XX
KW Chromosome mapping; human; hereditary disease; diagnosis; cancer;
KW hyperlipidemia; cardiovascular; neurodegenerative disorder; therapy;
KW autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
KW renal injury; amino aciduria; hypoglycaemia; male rat infertility;
KW hypertension; ss.
XX
OS Homo sapiens.
XX
PN W09940189-A2.
XX
PD 12-AUG-1999.
XX
PF 09-FEB-1999; 99WO-IB00282.
XX
PR 09-FEB-1998; 98US-0074121.
PR 13-APR-1998; 98US-0081563.
PR 10-AUG-1998; 98US-0096116.
PR 04-SEP-1998; 98US-0099273.
XX
PA (GEST ) GENSET.
XX
PI Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
DR WPI; 1999-600966/51.
DR P-PSDB; AAY59655.
XX
PT Extended cDNAs useful for expressing secreted proteins and to obtain
PT specific antibodies -
XX
PS Claim 1; Page 150; 244pp; English.
XX
CC This sequence encodes a human secreted protein of the invention. The
CC extended cDNAs (or genomic DNAs obtainable from them) may be used to
CC prepare PCR primers and probes. These are useful for forensic matching or
CC positive identification by DNA sequencing. They may also be used in
CC alternative fingerprint identification techniques. Antibodies against the
CC proteins encoded by the extended cDNAs are useful in identification of
CC tissue types or cell species, as well as identifying tissue specific
CC soluble proteins. The sequences can be used for chromosome mapping and
CC identification of genes associated with hereditary diseases or drug
CC response. signal sequences from the cDNAs can be used in construction of
CC secretion vectors. Other sequences derived from the extended cDNAs can be
CC used to clone upstream genomic DNA sequences including promoters. This is
CC in turn useful for identifying proteins that interact with promoter
CC sequences. Some of the proteins may be useful in diagnosing and treating
CC several disorders including, but not limited to: cancer, hyperlipidaemia,
CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and
CC rheumatic diseases, embryogenic disorders, hypertension, renal injury,
CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.
XX
SQ Sequence 497 BP; 122 A; 138 C; 90 G; 147 T; 0 other;

Query Match 99.2%; Score 476.4; DB 20; Length 497;
```

```
Best local Similarity 59.8%; Pred. No. 1.5e-108;
Matches 477; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ggctttgagcaatttttgcctgtgtcctcgtatcttccaggtccaccaccatgaagtcttta 60
Db 20 ggctttgagcaatttttgcctgtgtcctcgtatcttccaggtccaccaccatgaagtcttta 79

Qy 61 gcagtcctcggtaactctctgggagtttcccatcttcttggtctctgcccagaatccgacaaca 120
Db 80 gcagtcctcggtaactctctgggagtttcccatcttcttggtctctgcccagaatccgacaaca 139

Qy 121 gctgtccagctgacgcgtatccagctactgctcctgtctgtgatgaaagccctctgatgct 180
Db 140 gctgtccagctgacgcgtatccagctactgctcctgtctgtgatgaaagccctctgatgct 199

Qy 161 gaacccaactgctgtgcaacccaactgcgaccactgctgctcctaccactgcaacccaactgct 240
Db 200 gaacccaactgctgtgcaacccaactgcgaccactgctgctcctaccactgcaacccaactgct 259

Qy 241 gctttaccactgctcgttaaaagacacattccagttttaccacaaatgggttggggatctcccg 300
Db 260 gctttaccactgctcgttaaaagacacattccagttttaccacaaatgggttggggatctcccg 319

Qy 301 aaaggttagagtgctccctgagatggaatcagcttgagcttcttctgcaattggtcacaaact 360
Db 320 aaaggttagagtgctccctgagatggaatcagcttgagcttcttctgcaattggtcacaaact 379

Qy 361 attcatgctcctctgttatttctccaaactacttaacttgcctacagataatcccttttatct 420
Db 380 attcatgctcctctgttatttctccaaactacttaacttgcctacagataatcccttttatct 439

Qy 421 cttaatcagttattttcttctcaataaaaaataaactatgagcaacaaaaaaatcttcttct 478
Db 440 cttaatcagttattttcttctcaataaaaaataaactatgagcaacaaaaaaatcttcttct 497

RESULT 10
AAZ51765
ID AAZ51765 standard; DNA; 544 BP.
XX
AC AAZ51765;
XX
DI 01-JUN-2000 (first entry)
XX
DE Human breast specific gene clone Mam002.
XX
KW Human; breast specific gene; breast specific marker; BSG; diagnosis;
KW breast cancer; therapy; Mam002; ss.
XX
OS Homo sapiens.
XX
PN W0200008210-A1.
XX
PD 17-FEB-2000.
XX
PF 22-JUL-1999; 99WO-US16811.
XX
PR 04-AUG-1998; 98US-0095232.
XX
PA (DIAD-) DIADEXUS LLC.
XX
PI Sun Y, Recipon H, Cafferkey R;
XX
WPI; 2000-205737/18.
XX
CC Detecting, diagnosing, monitoring, staging, prognosticating, imaging
CC and treating breast cancer using protein product of breast specific
CC genes -
XX
PS Claim 7; Page 38; 45pp; English.
XX
CC This sequence represents a breast specific gene (BSG) clone. The BSG
CC sequences are also referred to as breast specific markers (BSM). The
```


CC invention relates to methods for diagnosing, monitoring, staging, imaging
CC and treating breast cancer (BC). The methods comprise measuring the
CC levels of BSG products in cells, tissues or body fluids of the patient
CC and comparing the measured levels of BSG, with BSG levels of a normal
CC human control. An antibody against the BSG sequences can be labelled and
CC used for imaging BC in a patient. The antibody can be conjugated to a
CC cytotoxic agent, and used for treating BC in a patient.

XX
SQ Sequence 544 BP; 142 A; 143 C; 94 G; 154 T; 11 other;

Query Match 98.7%; Score 473.8; DB 21; Length 544;
Best Local Similarity 99.0%; Pred. No. 6.8e-108;
Matches 475; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggtttgaagcatttttctgtgctccctgactcttcaggtcaccaccatgaagttcttta 60
DB |||||||
QY 36 ggtttgaagcatttttctgtgctccctgactcttcaggtcaccaccatgaagttcttta 95
DB |||||||
QY 61 gcagtcctggaactcttgaggatttccatcttcttggtctctgccagaaatccgacaaca 120
DB |||||||
QY 96 gcagtcctggaactcttgaggatttccatcttcttggtctctgccagaaatccgacaaca 155
DB |||||||
QY 121 gctgtccagtgacagctatccagctactgctctgtctgtgatgatgaagcccttgatgct 180
DB |||||||
QY 156 gctgtccagtgacagctatccagctactgctctgtctgtgatgatgaagcccttgatgct 215
DB |||||||
QY 181 gaaacactgctgtcgaacactgcgacactgctctgtctctaccactgcaaccacgct 240
DB |||||||
QY 216 gaaacactgctgtcgaacactgcgacactgctctgtctctaccactgcaaccacgct 275
DB |||||||
QY 241 gcttaccactgctgtcgaacactgcgacactgctctgtctctaccactgcaaccacgct 300
DB |||||||
QY 276 gcttaccactgctgtcgaacactgcgacactgctctgtctctaccactgcaaccacgct 335
DB |||||||
QY 301 aatggttagagtgctccctgagatgaatcagcttgatctctcgaattggtcacaact 360
DB |||||||
QY 336 aatggttagagtgctccctgagatgaatcagcttgatctctcgaattggtcacaact 395
DB |||||||
QY 361 attcatgctctctgatttcacacactacttaactcttgcctacgataatccctttatct 420
DB |||||||
QY 396 attcatgctctctgatttcacacactacttaactcttgcctacgataatccctttatct 455
DB |||||||
QY 421 ctaactcagttattttcttccaaataaaaaaataactatgacgaactaaanaaaaaa 480
DB |||||||
QY 456 ctaactcagttattttcttccaaataaaaaaataactatgacgaactaaanaaaaaa 515
DB |||||||

RESULT 11

AAZ65099

ID AAZ65099 standard; cDNA; 471 BP.

XX

AC AAZ65099;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PR 04-JUN-1998; 98US-0088021.
PR 04-JUN-1998; 98US-0088025.
PR 04-JUN-1998; 98US-0088028.
PR 04-JUN-1998; 98US-0088029.
PR 04-JUN-1998; 98US-0088030.
PR 04-JUN-1998; 98US-0088033.
PR 04-JUN-1998; 98US-0088326.
PR 05-JUN-1998; 98US-0088167.
PR 05-JUN-1998; 98US-0088202.
PR 05-JUN-1998; 98US-0088212.
PR 05-JUN-1998; 98US-0088217.
PR 05-JUN-1998; 98US-0088655.
PR 10-JUN-1998; 98US-0088722.
PR 10-JUN-1998; 98US-0088730.
PR 10-JUN-1998; 98US-0088734.
PR 10-JUN-1998; 98US-0088738.
PR 10-JUN-1998; 98US-0088740.
PR 10-JUN-1998; 98US-0088741.
PR 10-JUN-1998; 98US-0088742.
PR 10-JUN-1998; 98US-0088810.
PR 10-JUN-1998; 98US-0088811.
PR 10-JUN-1998; 98US-0088824.
PR 10-JUN-1998; 98US-0088825.
PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 12-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0089090.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089653.
PR 18-JUN-1998; 98US-0089601.
PR 18-JUN-1998; 98US-0089597.
PR 18-JUN-1998; 98US-0089598.
PR 19-JUN-1998; 98US-0089547.
PR 19-JUN-1998; 98US-0089548.
PR 19-JUN-1998; 98US-0089552.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 25-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 26-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.


```

XX PF 30-MAR-2000; 2000WO-US08439.
XX PR 02-JUN-1999; 99WO-US12252.
XX PR 23-JUN-1999; 99US-0141037.
PR 07-JUL-1999; 99US-0143048.
PR 20-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 28-JUL-1999; 99US-0146222.
PR 17-AUG-1999; 99US-0149396.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 08-OCT-1999; 99US-0158663.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28301.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.
PR 02-MAR-2000; 2000WO-US05004.
PR 15-MAR-2000; 2000WO-US05841.
PR 20-MAR-2000; 2000WO-US06884.
XX PF 20-MAR-2000; 2000WO-US07377.
PA (GETH ) GENENTECH INC.
XX PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Kijavini LJ, Napier MA, Pan J, Paoni NF;
PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;
PI Zhang Z;
XX PI WPI; 2001-032160/04.
DR P-PSDB; AAB65276.
XX PR PRO polynucleotides used to produce polypeptides used to target
PT bioactive molecules such as toxins, radiolabels or antibodies, to
PT specific cells, to cause targeted cell death -
XX PS Claim 2; Fig 281; 935pp; English.
XX CC The present invention describes human secreted and transmembrane PRO
CC proteins. The PRO proteins have cytosolic activity. The PRO proteins
CC can be used for targeted delivery of bioactive molecules, such as
CC toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
CC sequences, and their fragments, can be used as hybridisation probes, in
CC chromosomal and gene mapping, and in the generation of anti-sense RNA
CC and DNA. They may also be used to produce transgenic animals which are
CC used to develop and screen therapeutically useful reagents. The PRO
CC nucleotide and protein sequence can be used for tissue typing and in
CC treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
CC AAF44270 to AAF4470 represent PCR primers and hybridisation probes used
CC in the isolation of human PRO sequences. AAF44087 to AAF44265 and
CC AAB65154 to AAB65300 represent human PRO polynucleotide and protein
CC sequences given in the exemplification of the present invention.
XX SX Sequence 471 BP; 118 A; 131 C; 84 G; 138 T; 0 other;
Query Match 97.8%; Score 469.4; DB 22; Length 471;
Best Local Similarity 99.8%; Pred. No. 7.9e-107;
Matches 470; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 10 gcattttgtctgtgcctcgtatcttccaggtccaccacatgaagtcttagcagctcgt 69
DB 11 gattttttgtctgtgcctcgtatcttccaggtccaccacatgaagtcttagcagctcgt 69
QY 70 gtactcttggaggtttccattcttctgtctgtccaggaatccgacacagctgctcca 129
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
Db 61 gtactcttggaggtttccattcttctgtctgtccaggaatccgacacagctgctcca 120
QY 130 gctgacagctatccagctacttctgtctgtgatgatgaagcccttgatgctgaaacact 189
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
Db 121 gctgacagctatccagctacttctgtctgtgatgatgaagcccttgatgctgaaacact 180
QY 190 gctgtgcaaacactgcgcacactgctgctctctaccactgcaacacacgctctctacc 249
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
Db 181 gctgtgcaaacactgcgcacactgctgctctctaccactgcaacacacgctctctacc 240
QY 250 actgtctgtaaaagacattccagttttaccaccaatgggttggggatctcccgaaatgtaga 309
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
Db 241 actgtctgtaaaagacattccagttttaccaccaatgggttggggatctcccgaaatgtaga 300
QY 310 gctgtctccctgagatgggaatcagctgagctctctgcaattggtcacacactatcagct 369
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
Db 301 gctgtctccctgagatgggaatcagctgagctctctgcaattggtcacacactatcagct 360
QY 370 tctgtgatttccatccaaacttacccttgcctgacagatacccccttctctctaatcagct 429
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
Db 361 tctgtgatttccatccaaacttacccttgcctgacagatacccccttctctctaatcagct 420
QY 430 ttatttttttccaaataaaaaataactatgagcaacaaaaaa 480
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
Db 421 ttatttttttccaaataaaaaataactatgagcaacaaaaaa 471
RESULT 13
AAC97492
ID AAC97492 standard; cDNA; 471 BP.
XX AC AAC97492;
XX DT 28-FEB-2001 (first entry)
XX DE Human angiogenesis-associated protein PRO1160 cDNA, SEQ ID NO:159.
XX KW Human; angiogenesis-associated protein; PRO; endothelial cell growth;
KW cardiac hypertrophy; cardiovascular disorder; endothelial disorder;
KW angiogenic disorder; atherosclerosis; osteoporosis; hypertension;
KW myocardial infarction; diabetic retinopathy; rheumatoid arthritis;
KW Crohn's disease; psoriasis; endometriosis; ulcer; wound healing; cancer;
KW Alzheimer's disease; Huntington's disease; stroke; drug screening;
KW gene therapy; transgenic animal; ss.
XX OS Homo sapiens.
XX PN WO2000053753-A2.
XX PD 14-SEP-2000.
XX PF 05-JAN-2000; 2000WO-US00219.
XX PR 06-MAR-1999; 99WO-US05028.
PR 12-MAR-1999; 99US-0123957.
PR 14-MAY-1999; 99US-0134287.
PR 02-JUN-1999; 99WO-US12252.
PR 23-JUN-1999; 99US-0141037.
PR 20-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 01-SEP-1999; 99WO-US20111.
PR 08-SEP-1999; 99WO-US20594.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 05-OCT-1999; 99WO-US23089.
PR 30-NOV-1999; 99WO-US28313.
PR 30-NOV-1999; 99WO-US28409.
PR 02-DEC-1999; 99WO-US28564.
PR 02-DEC-1999; 99WO-US28565.
XX PA (GETH ) GENENTECH INC.
XX PI Ashkenazi AJ, Baker KP, Ferrara N, Gerber H, Goddard A;

```

Codowski PJ, Gurney AL, Hillan KJ, Kuo SS, Mark MR, Marsters SA;
 Paoni NF, Pittl RM, Williams PM, Wood WI;
 WPI: 2001-090793/10.
 P-PSDB; AAB53095.
 New isolated nucleic acid for producing a PRO polypeptide, analyzing
 genetic disorders and treating cardiovascular, endothelial or
 angiogenic disorders, such as atherosclerosis, wounds or cancer -
 Claim 58; Fig 63; 293pp; English.
 The invention relates to novel human angiogenesis-associated proteins
 designated PRO proteins (AAB53064-B53097), and to nucleic acids encoding
 PRO proteins. The invention also relates to vectors and host cells
 comprising a PRO nucleic acid, the recombinant production of a PRO
 protein, PRO antibodies specific for a PRO protein, fusion proteins
 comprising a PRO protein, agonists or antagonists of a PRO protein, and
 compounds which inhibit the expression of a PRO gene. The invention
 additionally encompasses methods of identifying modulators of PRO
 expression or activity; diagnosing a cardiovascular, endothelial or
 angiogenic disorder, or a susceptibility to such a disorder by detecting
 mutations in a PRO gene, or the expression level of a PRO gene within a
 particular tissue; treating a cardiovascular, endothelial or angiogenic
 disorder via the administration of a PRO protein, PRO nucleic acid, or
 PRO agonist or antagonist; a retroviral gene therapy vector comprising a
 PRO nucleic acid; and methods of inhibiting or stimulating endothelial
 cell growth, cardiac hypertrophy or PRO-induced angiogenesis via the
 administration of a PRO protein, or an agonist or antagonist thereof.
 PRO nucleic acids, PRO proteins, or an agonist or antagonist thereof.
 PRO agonists and PRO antagonists may be used as therapeutic agents to treat
 cardiovascular, endothelial or angiogenic disorders, such as
 atherosclerosis, osteoporosis, myocardial infarction, hypertension,
 diabetic retinopathy, rheumatoid arthritis, Crohn's disease, psoriasis,
 endometriosis, ulcers, wounds, cancer, Alzheimer's disease, Huntington's
 disease, or stroke. PRO nucleic acids are additionally useful in the
 recombinant production of PRO proteins, as hybridisation probes to
 screen libraries to isolate cDNAs with sequence identity to PRO proteins,
 to map genes encoding PRO proteins, to analyse genetic disorders, and in
 gene therapy. PRO nucleic acids can also be used to produce transgenic
 animals useful for the development and screening of potential
 therapeutic agents. The present sequence represents a cDNA encoding a PRO
 protein of the invention.

Sequence 471 BP: 118 A; 131 C; 84 G; 138 T; 0 other;

	Query Match	97.8%	Score 469.4	DB 22	Length 471
	Best Local Similarity	99.8%	Pred. No. 7.9e-107		
	Matches 470	Conservative 0	Mismatches 1	Indels 0	Gaps 0
Qy	10	gcattttgtctgtccctgattctcaggtcacccaccatgaagttcttagcagtcctg	65		
Dp	1	gcattttgtctgtccctgattctcaggtcacccaccatgaagttcttagcagtcctg	60		
Qy	70	gtactcttgggagtttccattcttctgtctctgtccacagaatccgacacagctgtccca	129		
Dp	61	gtactcttgggagtttccattcttctgtctctgtccacagaatccgacacagctgtccca	120		
Qy	130	gcgcacacgtatccagctactgctctgctatgatgaagccctctgctgtgaaacact	189		
Dp	121	gcgcacacgtatccagctactgctctgctatgatgaagccctctgctgtgaaacact	180		
Qy	190	gcctgtgcacacactgcgcaccactgcctcctaccactgcacccacgcgtctctacc	249		
Dp	181	gcctgtgcacacactgcgcaccactgcctcctaccactgcacccacgcgtctctacc	240		
Qy	250	actgctcgtaagacattccattttacccaaatgggttggggatctcccgaaatgtaga	309		
Dp	241	actgctcgtaagacattccattttacccaaatgggttggggatctcccgaaatgtaga	300		
Qy	310	gtgtgtcccttgagatggaaatcgcttgagctctctgcgaattggttcacaaactattcgt	369		

D6	301	ytgtgtcccttgagatggtatcagctctgagtctctctgtgcattggtgcacaaactattcaatgc	360
Qy	370	tctctgtgatttcacccaactactactctctggtccctacgataatcccccttattctctaatcagt	429
D6	361	tctctgtgatttcacccaactactactctggtccctacgataatcccccttattctctaatcagt	420
Qy	430	ttattttcttccaaacaaacaaacaaactatgagcaacaaacaaacaaacaaacaa	480
h	421	ttattttcttccaaacaaacaaacaaactatgagcaacaaacaaacaaacaaacaa	471

XX	AAAF44870	standard; cDNA; 488 BP.
XX	AAAF44870;	
XX	AAAF44870;	
XX	28-MAR-2001	(first entry)
XX	Human breast cancer	related protein coding sequence SEQ ID NO: 26.
XX	Human; breast cancer;	diagnosis; therapy; vaccine; ss.
XX	Homo sapiens.	
XX	WJ200078560-A2.	
XX	28-DEC-2000.	
XX	23-JUN-2000;	2000WO-US17536.
XX	23-JUN-1999;	59US-0410503.
XX	12-OCT-1999;	59US-0158980.
XX	(CORI-)	CORIXA CORP.
XX	Yuqiu J, Mitcham JI;	
XX	WPI; 2001-041426/05.	
XX	New polynucleotides	encoding breast tumor specific proteins, useful for
XX	prevention, treatment	and diagnosis of breast cancer -
XX	Claim 5; Page 127;	165pp: English.
XX	The present invention	provides the coding sequences for a number of
XX	breast cancer related	proteins. These can be used in vaccinations
XX	against, diagnosis of	and treatment of cancer, particularly breast
XX	cancer.	
XX	Sequence 488 BP;	135 A; 129 C; 84 G; 138 T; 2 other;

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2002, 17:16:38 ; Search time 6783.7 Seconds
(without alignments)
760.033 Million cell updates/sec

Title: US-09-248-178-59
Perfect score: 382
Sequence: 1 agcgggagcagcagctaaa.....attaaaaaaaaaaaaaaaaaa 382

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 2747244

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hct:*
9: gb_estl:*
10: gb_est2:*
11: gb_hct:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	382	100.0	400	9	AI302562
2	382	100.0	535	9	AA533894
3	382	100.0	699	10	BC939439
4	382	100.0	4676	11	BC023515
5	378.8	99.2	536	9	AI890693
6	377.2	98.7	436	9	AI377484
7	376.8	98.6	680	9	AI921707
8	375.8	98.2	567	9	BE046355
9	373.6	97.8	383	9	AA971201
10	373.6	97.5	394	9	AA703778
11	372.6	97.3	684	10	BF196380
12	371.8	97.3	425	9	AI091673
13	370.8	97.1	384	10	BF002805
14	370.4	97.0	627	10	BF445293
15	370.0	96.9	431	9	AA535981
16	370.0	96.9	466	9	AI928430
17	369.6	96.8	530	9	AW163074

C 15	369.2	96.6	446	9	AA451680
C 19	369.0	96.6	532	9	AW591085
C 20	367.8	96.3	609	9	AW338692
C 21	367.6	96.2	514	9	AW622108
C 22	366.8	96.0	416	9	AI675597
C 23	366.6	96.0	524	10	BF434100
C 24	366.6	96.0	791	9	AL514915
C 25	366.6	95.8	406	9	AI354236
C 26	366.6	95.8	583	9	AW673998
C 27	365.8	95.8	574	9	AW167217
C 28	365.8	95.8	624	10	BM049217
C 29	365.5	95.5	374	9	AI742813
C 30	365.5	95.5	384	9	AI195134
C 31	365.5	95.5	387	9	AI306585
C 32	365.5	95.5	388	9	AI538886
C 33	365.5	95.5	414	9	AI129909
C 34	365.5	95.5	455	9	AI669857
C 35	365.5	95.5	470	9	AI055643
C 36	365.5	95.5	471	9	AA722353
C 37	365.5	95.5	472	9	AI369794
C 38	365.5	95.5	473	9	AI144522
C 39	365.5	95.5	499	10	BG059808
C 40	365.5	95.5	532	9	AW572439
C 41	365.5	95.5	700	9	AI951963
C 42	364.8	95.3	570	9	AI934547
C 43	364.8	95.3	474	9	AI565670
C 44	364.8	95.3	550	10	BM142598
C 45	363.4	95.1	387	9	AA613497

ALIGNMENTS

RESULT 1
AI302562/c
LOCUS
DEFINITION qn52c02.x1 NCI_CGAP_Kid5 Homo sapiens cDNA clone IMAGE:1901858 3', mRNA sequence.
ACCESSION AI302562
VERSION AI302562.1 GI:3961908
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (Bases 1 to 400)
AUTHORS NCI-CoAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Unpublished (1997)
JOURNAL Tumor Gene Index
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgs@nci.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio-llnl.gov/bbrp/image/image.html
Insert length: 1055 Std Error: 0.00
Seq primer: -400P from Glbco
High quality sequence stop: 361.
Location/Qualifiers
1. 400
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1901858"
/clone_lib="NCI_CGAP_Kid5"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; 1st

SEQ ID NO: 59

strand cDNA was primed with a Not I - oligo(dT) primer (5'-AACTGGAAGATTCGGCGCAATATTTTITTTTTT 3'), double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 89 a 95 c 108 g 108 t

ORIGIN

Query Match 100.0%; Score 382; DB 9; Length 400;
Best Local Similarity 100.0%; Pred. No. 1.4e-67;
Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agcgggagcagaagctaaagccaaagcccaagagtggtgagtgccagcactgggtgccca 60
|||||
Db 383 AGCGGGAGCAGAGCTAAAGCCAAAGCCCAAGAGAGTGCCAGTGCCAGCCTGGTGCCA 324
|||||

QY 61 gtaccagtcaccataaacagtgccagtgccagtgccagcaccagtggtggttcagtgctg 120
|||||
Db 323 GTACCACTACCAATAACAGTGCCAGTGCCAGTGCCAGTGCCAGCCTGGTGCCA 264
|||||

QY 121 gtaccagtcagcagccactctcacatttggtggtctctgctggccttggtgagctggcg 180
|||||
Db 263 GTGCCAGCTGACCGCCACTCTCACATTTGGGCTCTTCGCTGCCCTTGGTGAGCTGGTG 204
|||||

QY 181 ccagcaccagtgccagctctggtgctggtgttctctccacaaagtggagatttttagatatt 240
|||||
Db 203 CCAGCACCAGTGCCAGCTGCTGGTGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 144
|||||

QY 241 gtaatactccagctcttctctccaaagcaggggtgcatactccagaaactcctcaaacac 300
|||||
Db 143 GTTAATCCTCCAGCTCTTCTCTCTCAAGCCAGGTCATCTCTCAGAAACCTTACTCAACAC 84
|||||

QY 301 agcactctaggcagccactatcaatcaattgaagtggacactctgcattaaatctatttg 360
|||||
Db 83 AGCACTCTAGCGAGCCACTATCAATCAATTGAAGTTGACACTCTGCAATTAATCTATTTC 24
|||||

QY 361 ccattaaaaaiaaaaaaaaaa 382
|||||
Db 23 CCATTAAAAAIAAAAAAAAAA 2

RESULT 2
AA535894/c
LOCUS
DEFINITION
n94a08.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:927542 3', similar to contains element TAR1 repetitive element ;, mRNA sequence.
ACCESSION
AA535894
VERSION
AA535894.1 GI:2280147
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 535)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 584 Std Error: 0.00

Seq primer: -t0m13 fwd. ET from Amersham
High quality sequence stop: 269.

FEATURES
Source

Location/Qualifiers
1..535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:927542"
/clone_lib="NCI_CGAP_Co3"
/sex="pooled"
/tissue_type="colon"
/lav_host="DH108"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization."

BASE COUNT 112 a 129 c 150 g 144 t

Query Match 100.0%; Score 382; DB 9; Length 535;
Best Local Similarity 100.0%; Pred. No. 1.3e-67;
Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agcgggagcagaagctaaagccaaagcccaagagtggtgagtgccagcactgggtgccca 60
|||||
Db 386 AGCGGGAGCAGAGCTAAAGCCAAAGCCCAAGAGAGTGCCAGTGCCAGCCTGGTGCCA 327
|||||

QY 61 gtaccagtcaccataaacagtgccagtgccagtgccagcaccagtggtggttcagtgctg 120
|||||
Db 326 GTACCACTACCAATAACAGTGCCAGTGCCAGTGCCAGTGCCAGTGCTTCAGTGCTG 267
|||||

QY 121 gtaccagtcagcagccactctcacatttggtggtctctgctggtggtggtggtggtg 180
|||||
Db 266 GTGCCAGCTGACCGCCACTCTCACATTTGGGCTCTTCGCTGCCCTTGGTGAGCTGGTG 207
|||||

QY 181 ccagcaccagtggtggtctctggtggtggttctctccacaaagtggagatttttagatatt 240
|||||
Db 206 CCAGCACCAGTGCCAGCTGCTGGTGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 147
|||||

QY 241 gtaatactccagctcttctctccaaagcaggggtgcatactccagaaactcctcaaacac 300
|||||
Db 146 GTTAATCCTCCAGCTCTTCTCTCTCAAGCCAGGTCATCTCTCAGAAACCTTACTCAACAC 87
|||||

QY 301 agcactctaggcagccactatcaatcaattgaagtggacactctgcattaaatctatttg 360
|||||
Db 86 AGCACTCTAGCGAGCCACTATCAATCAATTGAAGTTGACACTCTGCAATTAATCTATTTC 27
|||||

QY 361 ccattaaaaaiaaaaaaaaaa 382
|||||
Db 26 CCATTAAAAAIAAAAAAAAAA 5

RESULT 3
BG939439/c
LOCUS
DEFINITION
cr54gl1.x1 J1a bone marrow stroma Homo sapiens cDNA clone HBN5C_cr54gl1 3', mRNA sequence.
ACCESSION
BG939439
VERSION
BG939439.1 GI:14338811
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 699)
Jia, L.B., You, M.F., Touchman, J.W., Bouffard, G.G., Beckstrom-Sterner, S.W., Green, E.D., Powell, J.L., Yang, L.M., Roray, P.G., Hotchkiss, R.N. and Francomano, C.A.
SGAP: The Skeletal Genome Anatomy Project

JOURNAL COMMENT	REFERENCE AUTHORS TITLE JOURNAL
Unpublished (1997) Contact: Libin Jia Medical Genetics Branch National Human Genome Research Institute 10/10C101, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA Tel.: 301-402-4877	

BASE COUNT	145 a	187 c	181 g	186 t
ORIGIN				
Query Match:	100.0%	Score 382;	DB 10;	Length 699;

RESULT	4
LOCUS	BC023515
DEFINITION	Homo sapiens, clone IMAGE:4135681, mRNA.
ACCESSION	BC023515
VERSION	BC023515.1 GI:18497333
KEYWORDS	HTC.
SOURCE	human.
ORGANISM	Homo sapiens
LOCUS	BC023515
DEFINITION	Homo sapiens, clone IMAGE:4135681, mRNA.
ACCESSION	BC023515
VERSION	BC023515.1 GI:18497333
KEYWORDS	HTC.
SOURCE	human.
ORGANISM	Homo sapiens

Eukaryota: Mollazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
1 (bases 1 to 4676)
Strausberg, R.
Direct Submission
Submitted (07-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgaptc@mail.nih.gov
Tissue Procurement: ATCC
cDNA library Preparation: Rubin Laboratory
cDNA library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mgc@nih.gov
Shenchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, T.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Madur, S.Q.L., Nasello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Standrip, S., Thomas, P.J.,
Tiongson, E., Touchman, J.W., Tsurgou, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

BASE COUNT	1217 a	1243 c	1238 g	978 t
ORIGIN				

[illegible]

QY	241	gttaatcttcgcagtgcttctctcctaagccagggtgcatcctctcagaataacctactcaaac	300
DB	142	TTTAACTCCTGCCAGTCTTTCITCTCAAGCGAGGTGCATCCICAGAAACCTACTCAAC	83
QY	301	aycacctctaggcgaccaciatcatcaattgaagttagcactctgcattaaatctattg	360
DB	82	TT	23
QY	361	ccattaaaaaaaataaaaaaaaaa	382
DB	22	CCATTAAAAAATAAAAAAAAAA	1
 RESULT 6 AI377484/c			
LOCUS			
DEFINITION			
AI377484 tc37b03.xl Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone IMAGE:2066765 3' similar to contains element LTR10 repetitive element ; , mRNA sequence.			
ACCESSION AI377484 GI:4187337			
VERSION AI377484.1			
KEYWORDS EST.			
SOURCE Human.			
ORGANISM Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE 1 (bases 1 to 436)			
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.			
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index			
JOURNAL Unpublished (1997)			
COMMENT Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 735 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 400.			
FEATURES			
Source			
1..436			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/clone="IMAGE:2066765"			
/clone_lib="Soares_total_fetus_Nb2HF8_9w"			
/dev_stage="8-9 weeks"			
/lab_host="DH10B"			
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dI) primer [5', TTTTACCAATCTGAAGTGGGAGCGCGCTTAATTTTTFITTTT 3'] Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."			
BASE COUNT 97 a 105 c 118 g 116 t			
ORIGIN			
Query Match 98.7%; Score 377.2; Db %; Length 436;			
Best local Similarity 99.2%; Pred. No. 13e-66;			
Matches 379; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
QY	1	agggcgagcagaagctaaagccccaaagcccagaagtgccagtcgccagcactggtgcc	60
DB	385	AGGCGGAGCAGAGAAGCTAAAGCCAAAGCCCAAGAGATGGCAGTGCCAGCACTGGTCCA	326
QY	61	gtaccagtaccaataacagtcgcagtcgcagtcgccagccagtcggtggttcagtcgtg	120
DB	325	GTACCAGTACCAATAACAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC	266
QY	121	gtgccagctgcagcgcacactctcacattttgggtctcttcgcttggcttgggtgagctg	180

QY	361	ccattataaaaaaaaaaaaaa 382
Db	4651	CCATTAAAAAAAAAAAAAAAAA 4672
RESULT	5	
AI890693/C		
LOCUS		
DEFINITION		AI890693 536 bp mRNA linear EST 07-MAR-2000
ACCESSION		wm92d06.xl NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2443403
VERSION		similar to TR:076058 076058 DJ1409.2 ;contains element MER4
KEYWORDS		repetitive element ;, mRNA sequence.
SOURCE		AI890693 1 GI:5595857
ORGANISM		human.
REFERENCE		Homo Sapiens
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL		1 (bases 1 to 536)
COMMENT		NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
		National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
		Tumor Gene Index
		Unpublished (1997)
		Contact: Robert Strausberg, Ph.D.
		Email: cgapbs@mail.nih.gov
		Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
		Emmert-Buck, M.D., Ph.D.
		cDNA Library Preparation: Life Technologies, Inc.
		cDNA Library Arrayed by: Greg Lennon, Ph.D.
		DNA Sequencing by: Washington University Genome Sequencing Center
		Clone distribution: NCI-CGAP clone distribution information can be
		found through the I.M.A.G.E. Consortium/LLNL at:
		www-bio.llnl.gov/dbbrp/image/image.html
		Insert Length: 889 Std Error: 0.00
		Seq primer: -400P from Gibco
		High quality sequence stop: 391.
FEATURES		
source		Location/Qualifiers
		1..536
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone="IMAGE:2443403"
		/clone_lib="NCI_CGAP_Ut2"
		/tissue_type="moderately-differentiated endometrial
		adenocarcinoma 3 pooled tumors"
		/lab_host="DH10B"
		/note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: SalI;
		Site:2: NotI; Cloned unidirectionally. Primer: Oligo di.
		Average insert size 1.85 kb. Life Technologies catalog #:
		11539-012"
BASE COUNT		116 a 139 c 140 g 141 t
ORIGIN		
Query Match		99.2%; Score 378.8; DB 9; Length 536;
Best Local Similarity		99.5%; Pred. No. 5.8e-67;
Matches		380; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	1	agcgggagcagaagctaaagccaaagccaaagagtagtgcagtcgacgacctggcgcca 60
Db	382	AGCGGGAGCAGAAGCTAAGCCAAAGCCAAAGCCAGAGTGCAGTGCACGACCTGGTGCCA 323
QY	61	gtaccagttaccaataacagtgcagtcgacgtgccagtcgacgaccagtggtggttcagtgctg 120
Db	322	GTACCAAGTACCAATAACAGTGCAGTGCAGTGCAGTGCACGACCTGGTGCCA 263
QY	121	gtgccagcctgaccgcacactctcacatttgggctctcttcgctggccttggtggagctggtg 180
Db	262	GTGCCAGCTGACGGCCACTCTCACATTTCGGGCTCTTCGCTGGCCCTGGTGGAGCTGGTG 203
QY	181	ccagaccagtggcagctctggtgcctgtggtttctctccacaaagtgcagatttcagatatt 240
Db	202	CCAGCACAGTGGCAGTTCCTGGTGCCTGTGGTTTCCTCCACAAAGTCAGATTTCAGATAAT 143

Db	265	GTGCCAGCCCTGACCGGCACACTCTCACATTTGGGCTCTTCGCTGGCCITGGTGGAGCIGGIG	
Qy	181	ccagcacacagtgcagctcgtggtcgctctggcttcttcctcaagaatgagatttttagatat	
Db	205	CCAGCACCACTGGCAGCTTTGGTGSCCTGTGGTTTCTCTCTACAAGTGAGATTIITAGATAIT	
Qy	241	gttaactcctgcagtgctcttctctccaagccagggtgcatactcagaaaacctactcaaac	
Db	145	GTTAATCTCGCAGTCTTTCTTCTTCTTAAGCCAGGGTGATCTCTCAGAACCCTACTCAACAC	
Qy	301	agcaactagcgagccactcatcaatcaattgaagtgcacctctgcattaataatctattg	
Db	85	AGCACTCTAGCAGCCACTATCATCAATTCAGATTGACACTCTGCATTTAAATCTAIITG	
Qy	361	ccattaaaaaaaaaaaaaaa	382
Db	25	CCATTTCAAAAAAAAAAAAAA	4

RESULT 7
 A1921707/c
 LOCUS
 DEFINITION w029db10.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2456731 3' similar to TR:076058 O76058 DJ1409.2 ;contains element LTR5 repetitive element ; , mRNA sequence.
 ACCESSION A1921707
 VERSION A1921707.1 GI:5657671
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 680)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/hci/cgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:
www-bio.llnl.gov/bbrp/image/image.html
 Insert length: 1142 Std Error: 0.00
 Seq primer: -400P from Gibco
 High quality sequence stop: 423.
 Location/Qualifiers
 1..680
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2456731"
 /tissue_type="poorly differentiated adenocarcinoma with signed ring cell features"
 /lab_host="DH10B"
 /note="Organ: stomach; Vector: pCMV-SPORT6; Site:1; Salt: dl; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT." Average insert size 1.69 kb. Life Technologies catalog #: 11349-011"

BASE COUNT	143 a	184 c	176 g	176 t	1 others
ORIGIN					

Query Match 98.6% Score 376.8; DB 9; Length 680;
 Best Local Similarity 99.5%; Pred. No. 1.4e-66;
 Matches 378; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1	aggcgggagcagaagctaaagccaaagccccgaagagtgccagtcgccagcactggtgccca	60
----	---	--	----


```

/clone="IMAGE:1140976"
/clone_lib="Stratagene hNT neuron (#937233)"
/dev_stage="hNT neurons"
/lab_host="SOLR (kanamycin resistant)"
/notes="Vector: pBluescript SK-; Site.1: EcoRI; Site.2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Insert differentiated, post mitotic hNT neurons. Average size: 1.5 kb; Uni-ZAP XR vector; -5' adaptor sequence: 5' GAATTCGGACGAC 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."
BASE COUNT 91 a 94 c 109 g 100 t
ORIGIN

Query Match 97.5%; Score 372.6; DB 9; Length 394;
Best Local Similarity 98.9%; Pred. No. 1.le-65;
Matches 375; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 aggcggagcagaagctaaagccaaagcccaagagagtgccagtcgagcactggcgcca 60
|||||
Db 379 AGCGGGAGCAGACGCTTAAGCCAAAGCCCAAGAGAGTGCCAGTGGTGGCTTCAGTGC 320

Qy 61 gtaccagtcaccataaagcagtgccagtgccagtcgagcagcagtcgagtcgagtcg 120
|||||
Db 319 GTACCAAGTACCAATACAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 260

Qy 121 gtgcagcctgagcgcagcactcactcactcactcactcactcactcactcactcact 180
|||||
Db 259 GTGCCAGCTGTGAGCCGACCTCACAATTTGGGCTCTTCGCTGGCTTGGTGGAGCTGG 200

Qy 181 ccagcagcagtcgagcagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgag 240
|||||
Db 199 CCAGCAGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 140

Qy 241 gttactctgcagtccttctcctcagcagagtgatctcagcagcagcagcagcagcag 300
|||||
Db 139 GTTAATCTGCGAGTCCTTCTCAAGCCAGGTCATCTCAGAAACCTACACAC 80

Qy 301 agcactctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 360
|||||
Db 79 AGCACTCTAGCAGCAGCAGTATCAATCAATTAAGTTGAGTTGACACCTCGCATTAAT 20

Qy 361 ccattataaataaataaataa 379
|||||
Db 19 CCATTTCTGAAAAAATAA 1

RESULT 11
BF196380/c 684 bp mRNA linear EST 03-NOV-2000
LOCUS 7n69d03.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3569717 3'
DEFINITION similar to TR:076058 076058 DJ1409.2 ;contains element MER1
repetitive element ; mRNA sequence.
ACCESSION BF196380
VERSION BF196380.1 GI:11084264
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 684)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, Ite
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center

```

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: i:fo:image.llnl.gov

High quality sequence stop: 389.

Location/Qualifiers

source

1..684

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3569717"

/clone_lib="NCI_CGAP_Ov18"

/tissue_type="fibrothoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: ovary; Vector: pi713p-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCGACATTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pi713 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 151 a 179 c 176 g 177 t

ORIGIN

Query Match 97.5%; Score 372.6; DB 10; Length 684;

Best Local Similarity 98.9%; Pred. No. 9.8e-66;

Matches 375; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 aggcggagcagaagctaaagccaaagcccaagagagtgccagtcgagcactggcgcca 60

|||||

Db 379 AGCGGGAGCAGACGCTTAAGCCAAAGCCCAAGAGAGTGCCAGTGGTGGCTTCAGTGC 320

Qy 61 gtaccagtcaccataaagcagtgccagtgccagtcgagcagcagtcgagtcgagtcg 120

|||||

Db 319 GTACCAAGTACCAATACAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 260

Qy 121 gtgcagcctgagcgcagcactcactcactcactcactcactcactcactcactcact 180

|||||

Db 259 GTGCCAGCTGTGAGCCGACCTCACAATTTGGGCTCTTCGCTGGCTTGGTGGAGCTGG 200

Qy 181 ccagcagcagtcgagcagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgag 240

|||||

Db 199 CCAGCAGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 140

Qy 241 gttactctgcagtccttctcctcagcagagtgatctcagcagcagcagcagcagcag 300

|||||

Db 139 GTTAATCTGCGAGTCCTTCTCAAGCCAGGTCATCTCAGAAACCTACACAC 80

Qy 301 agcactctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 360

|||||

Db 79 AGCACTCTAGCAGCAGCAGTATCAATCAATTAAGTTGAGTTGACACCTCGCATTAAT 20

Qy 361 ccattataaataaataaataa 379

|||||

Db 19 CCATTTCTGAAAAAATAA 1

RESULT 12

BF196380/c 425 bp mRNA linear EST 13-FEB-1999

LOCUS 0023e09.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone

DEFINITION IMAGE:1567048 3', mRNA sequence.

ACCESSION A1091673

VERSION A1091673

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 425)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, Ite

I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center

Matches 381;		Conservative	0;	Mismatches	0;	Indels	1;	Gaps	1;
Qy	1	agcgggagcagaagctaaagccaaagcccaagagtggtgcaagtgccagcactggtgcca 60							
Db	385	AGCGGGAGGAGAGCTAAAGCCAAAGCCCAAGAGAGTGCCAGTCCAGCAGCTGCTGCCA 326							
Qy	61	gtaccagtaccataaacagtgccagtgccagtgccagcaccagtggtggcttcagtgctg 120							
Db	375	GTACCACTACCAATAACAGTGCCAGTGCCAGTGCCAGCACCAGTGGTGGCTTCAGTGCTG 266							
Qy	121	gtgcagcctgaccgccaactctcaacatttggtctctcgtggtggtggtggtggtg 180							
Db	265	GTGCCAGCCTGACCGCCACTCTCACATTGGGCTCTTCGCTGGCTTGGTGAGCT-GTG 207							
Qy	181	ccagcaccagtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 240							
Db	206	CCAGCACCAGTGCGAGCTCTGGTGGCTGTGGTTCTCTCTCTCTCTCTCTCTCTCTCT 147							
Qy	241	gttaatectgccaagtcttcttcttcttcttcttcttcttcttcttcttcttcttctt 300							
Db	146	GTTAATCTCTGCCAGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 87							
Qy	301	agcactctaggcagccactatcaatcaatgaaattgacactctgcatataatctatttg 360							
Db	86	AGCACTCTAGGCGCCACTATCAATCAATTGAAGTTGACACTCTGCAATTAATCTATTG 27							
Qy	361	ccattaaaaaa 382							
Db	26	CCATTAAAAA 5							

Search completed: April 29, 2002, 17:16:41
Job time: 13200 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2002, 17:17:18 ; Search time 6783.7 seconds

(without alignments)

1144.029 Million cell updates/sec

Perfect score: 575

Sequence: 1 accatataataaaaggatat.....aaaggaaaaa 575

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hic:*

9: gb_estl:*

10: gb_estc2:*

11: gb_hic:*

12: gb_gss:*

13: em_gss_hum:*

14: em_gss_inv:*

15: em_gss_pin:*

16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	522.2	90.8	947	9	AL133916
2	519	90.3	757	9	AW955684
3	504	87.7	608	9	AW950828
4	498.6	86.7	549	9	AI346341
5	497	86.4	541	9	AI867454
6	488	84.9	532	10	BM263376
7	423	73.6	467	10	BF432231
8	410	71.3	454	9	AI655270
9	410	71.3	455	9	AI421279
10	405	70.4	449	10	BF439949
11	402	69.9	454	9	AW014882
12	397.6	69.1	486	10	N54784
13	384.8	66.9	432	9	AA775552
14	366	63.7	410	9	AA626243
15	336	58.4	380	9	AI341407
16	334.2	58.1	423	10	N62351
17	328.4	57.1	409	10	N59253

18	279.6	48.6	422	9	BE175639
19	252	43.3	404	10	BF114841
20	250	43.3	294	9	AA456968
21	241	41.9	324	9	AA364013
22	237.8	41.4	430	9	AI743400
23	233	40.5	277	9	AA457077
24	232	40.3	276	9	AI358918
25	214.8	37.4	463	10	N76721
26	190	15.7	430	9	AI743400
27	88.4	15.4	415	9	BE102996
28	88.2	15.3	352	9	AI577685
29	88.2	15.3	360	9	AW534413
30	72.6	12.7	273	9	AA585385
31	56.8	9.5	57	10	BC231440
32	44	7.7	200	9	AA998336
33	43.6	7.6	1101	12	CNS00EYP
34	43	7.5	487	10	BM353500
35	42.8	7.4	517	10	C90235
36	42.6	7.4	386	10	BI276354
37	42.4	7.1	536	12	CNS00L44
38	41.2	7.2	334	9	AA714041
39	41.2	7.2	455	10	C92323
40	41.2	7.2	758	9	AL666548
41	40.8	7.1	429	10	C84124
42	40.6	7.1	394	10	BF457592
43	40.6	7.1	1067	12	CNS014QG
44	40.4	7.0	364	10	C92261
45	40.4	7.0	868	12	CNS010YV

ALIGNMENTS

RESULT 1

AL133916

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL133916
DKFZp761N0114.1 761 (synonym: hamy2) Homo sapiens cDNA clone
DKFZp761N0114.5', mRNA sequence.
947 bp mRNA linear EST 25-FEB-2000

AL133916.1 GI:6602103

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 947)

Ansorge, D., Winkler, U., Mewes, H.W., Weil, B. and Wiemann, S.)

EST (Ansorge, D., Winkler, U., Mewes, H.W., Weil, B. and Wiemann, S.)

Unpublished (1999)

Contact: Ansorge W

MIPS

Am Klopiergasse 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;

Sequenced by EMBL (European Molecular Biology Laboratory);

Heidelberg/Germany) within the cDNA sequencing consortium of the

German Genome Project.

No sl sequence available

This clone (DKFZp761N0114) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

Location/Qualifiers

1..947

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="DKFZp761N0114"

/clone_lib="761 (synonym: hamy2)"

/cissue_type="amygdala"

/dev_stage="adult"

/lab_host="DH10B"

/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI"

313 170 C 182 G 282 T

BASE COUNT

SEQ ID NO: 65

27472414

27472414

27472414

FEATURES		Seq primer: Reverse.		Location/Qualifiers	
source		1..757			
		/organism="Homo sapiens"			
		/db_xref="taxon:9606"			
		/clone_lib="MAGE resequences, MAGD"			
		/note="Vector: pBluescriptSKm"			
BASE COUNT		240 a		146 g	
ORIGIN		210 t			
		Query Match		50.3%; Score 519; DB 9; Length 757;	
		Best Local Similarity		98.4%; Pred. No. 3.4e-93;	
		Matches 566; Conservative		0; Mismatches 5; Indels 4; Gaps 4;	
QY	1	actgataataaaggatattccataatgaattttttatactgcattctttacattacattaccca	60		
DB	59	acttgatataaaaggatattccataatgaattttttatactgcattctttacattaccca	118		
QY	61	ctaaatacgtttattgcttgatgaagacctttccacagaatccctatggtgacgatttca	120		
DB	119	ctaaatacgtttattgcttgatgaagacctttccacagaatccctatggtgacgatttca	178		
QY	121	cttggtacttcataccatgctttaaagggcgagtttctcaaaaagcagaacacatgcc	180		
DB	179	cttggtacttcataccatgctttaaagggcgagtttctcaaaaagcagaacacatgcc	238		
QY	181	gccagttctcaagtcttctctcaactccatttgaatglaaaggcgctggcccccaatg	240		
DB	239	gccagttctcaagtcttctctcaactccatttgaatglaaaggcgctggcccccaatg	298		
QY	241	tgaggaggtccgaacattttctgaattccattttcttctgcgctgaatgacagttt	300		
DB	299	tgaggaggtccgaacattttctgaattccattttcttctgcgctgaatgacagttt	358		
QY	301	ctgtcattactgagttcccgatctttcccaagggtgtgtgattttcaaaaggccagcta	360		
DB	359	ctgtcattactgagttcccgatctttcccaagggtgtgtgattttcaaaaggccagcta	417		
QY	361	atagccagaaatcatgacctgaagagagatgaaatttcaagctgagccagcagga	420		
DB	418	atagccagaaatcatgacctgaagagagatgaaatttcaagctgagccagcagga	475		
QY	421	gtccagatgagcgaaggtcttctgagaatcagccatttggtaaaaaagatttttaag	480		
DB	476	gtccagatgagcgaaggtcttctgagaatcagccatttggtaaaaaagatttttaag	534		
QY	481	ctttatgtttatccatgagccatagaaaggtatggtattgttttaagaactattttaa	540		
DB	535	ctttatgtttatccatgagccatagaaaggtatggtattgttttaagaactattttaa	594		
QY	541	gtgtccagaccccaaaaggaaaaaataaaaaa	575		
DB	595	gtgtccagaccccaaaaggaaaaaataaaaaa	629		
RESULT 3		AW950828/c		608 bp	
LOCUS		AW950828		linear	
DEFINITION		EST362856 MAGE resequences, MAGA Homo sapiens cDNA, mRNA sequence.		EST 01-JUN-2000	
ACCESSION		AW950828			
VERSION		AW950828.1		GI:8140485	
KEYWORDS		EST.			
SOURCE		human.			
ORGANISM		Homo sapiens			
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
TITLE		1 (bases 1 to 608)			
JOURNAL		Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C., Holt			
COMMENT		I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and			
		Quackenbush, J.			
		Assessment of gene expression patterns in a model of colon tumor			
		metastasis using a 19,200 element cDNA microarray			

ORIGIN		Query Match		90.8%; Score 522.2; DB 9; Length 947;	
		Best Local Similarity		98.8%; Pred. No. 7.4e-94;	
		Matches 568; Conservative		0; Mismatches 3; Indels 4; Gaps 4;	
QY	1	actgataataaaggatattccataatgaattttttatactgcattctttacattaccca	60		
DB	241	acttgatataaaaggatattccataatgaattttttatactgcattctttacattaccca	300		
QY	61	ctaaatacgtttattgcttgatgaagacctttccacagaatccctatggtgacgatttca	120		
DB	301	ctaaatacgtttattgcttgatgaagacctttccacagaatccctatggtgacgatttca	360		
QY	121	cttggtacttcataccatgctttaaagggcgagtttctcaaaaagcagaacacatgcc	180		
DB	361	cttggtacttcataccatgctttaaagggcgagtttctcaaaaagcagaacacatgcc	420		
QY	181	gccagttctcaagtcttctctcaactccatttgaatglaaaggcgctggcccccaatg	240		
DB	421	gccagttctcaagtcttctctcaactccatttgaatglaaaggcgctggcccccaatg	480		
QY	241	tgaggaggtccgaacattttctgaattccattttcttctgcgctgaatgacagttt	300		
DB	481	tgaggaggtccgaacattttctgaattccattttcttctgcgctgaatgacagttt	540		
QY	301	ctgtcattactgagttcccgatctttcccaagggtgtgtgattttcaaaaggccagcta	360		
DB	541	ctgtcattactgagttcccgatctttcccaagggtgtgtgattttcaaaaggccagcta	595		
QY	361	atagccagaaatcatgacctgaagagagatgaaatttcaagctgagccagcagga	420		
DB	600	atagccagaaatcatgacctgaagagagatgaaatttcaagctgagccagcagga	657		
QY	421	gtccagatgagcgaaggtcttctgagaatcagccatttggtaaaaaagatttttaag	480		
DB	658	gtccagatgagcgaaggtcttctgagaatcagccatttggtaaaaaagatttttaag	716		
QY	481	ctttatgtttatccatgagccatagaaaggtatggtattgttttaagaactattttaa	540		
DB	717	ctttatgtttatccatgagccatagaaaggtatggtattgttttaagaactattttaa	776		
QY	541	gtgtccagaccccaaaaggaaaaaataaaaaa	575		
DB	777	gtgtccagaccccaaaaggaaaaaataaaaaa	811		
RESULT 2		AW955684		757 bp	
LOCUS		EST367754		linear	
DEFINITION		MAGE resequences, MAGD Homo sapiens cDNA, mRNA sequence.		EST 01-JUN-2000	
ACCESSION		AW955684			
VERSION		AW955684.1		GI:8145367	
KEYWORDS		EST.			
SOURCE		human.			
ORGANISM		Homo sapiens			
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
TITLE		1 (bases 1 to 757)			
JOURNAL		Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C., Holt			
COMMENT		I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and			
		Quackenbush, J.			
		Assessment of gene expression patterns in a model of colon tumor			
		metastasis using a 19,200 element cDNA microarray			
		Unpublished (2000)			
		Contact: John Quackenbush			
		The Institute for Genomic Research			
		9712 Medical Center Dr., Rockville, MD 20850, USA			
		Tel: 301 838 3528			
		Fax: 301 838 0208			
		Email: johnq@tigr.org			
		Plate: 82			

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 549)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap/>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Eickert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 550 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 446.
Location/Qualifiers
Source 1..549
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1926425"
/clone_lib="NCI-CGAP_C08"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT730-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from a
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaudo.
BASE COUNT 160 a 107 c 110 g 172 t
ORIGIN
Query Match 56.7%; Score 498.6; DB 5; Length 549;
Best Local Similarity 98.6%; Pred. No. 4.1e-83;
Matches 549; Conservative 0; Mismatches 4; Indels 4; Gaps 4;
QY 14 aggtatccataatgaatattttactgacatcccttgcagcattcttggctacttca 73
DB 549 AGGATAATCCATAAATGAAATATTTTATCTGCAATCCTTTACATAGCCCACTAAATCGTAT 490
QY 74 tgcctgtagagcctttccacagagatcctatggattgcagcattcttggctacttca 133
DB 489 TCGTTGATGAAGACCTTTTCACAGAAATCCTATGGATTCGAGGATTTTCACTTGGCTACTICA 430
QY 134 taccatgccttaagaggggcagtttcttcaaaagcagaaacatgcgcagttctcaag 193
DB 429 TACCCATGCCITAAAGAGGGGCAGTTTCTCAAAAGCAGAAACATGCCGCCAGTCTTCAAG 370
QY 164 ttctctcttaactccatttgaatgtaagggcagctgcccccaatgtggggaggctccga 253
DB 369 TTTTCCTCTTAACITCCATTTTGAATGTAAAGGCAGCTGGCCCCCAATGTGGGAGGTCCGA 310
QY 254 acattttctgaattcccatcttctgttcgcggctaaatgcagtttctgtcattactta 313
DB 309 ACAHTTTCGAATTCGCCATTTCTTGTCGGCGTAAATGACAGTTTCTGTCAITACTTA 250
QY 314 gattccccgattttccccaaaggtgttgattttacaaagggccagctaatagccagaaatc 373
DB 249 GATT-CCGATCTTTCCCAAAGGTGTTGATTTTACAAAGAGGCCAGCTTAAG-CAGAAATC 192
QY 374 atgacctgaaagagagatgaaatttcaagctgtgagccagcagggagctccagtaggc 433
DB 191 ATGACCCCTGAAGAGAGAGAGAGAA-TTCAAGCTGTGAGCCAGCCAGGAGCT-CAGTATGCC 134

JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 19
Seq primer: Forward.
Location/Qualifiers
Source 1..608
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:resesquences_MAGA"
/note="Vector: paluescriptskm"
BASE COUNT 184 a 115 c 123 g 186 t
ORIGIN
Query Match 87.7%; Score 504; DB 9; Length 608;
Best Local Similarity 99.3%; Pred. No. 3.4e-90;
Matches 548; Conservative 0; Mismatches 0; Indels 4; Gaps 4;
QY 1 acctgataaaagatccataatgaatttttactgcatccctttacattagcca 50
DB 548 ACTTGATATAAAGAGATATCCATATGATTTTATCTGCAATCCTTTACATAGCCA 489
QY 61 ctataatcgtttatgctgatgaagaccctttccacgaatcctatgattgacagattca 120
DB 488 CTAATACGTTATTGCTTGATGAAGACCTTTCACAGAAATCCTATGATTCGAGCAATTC 429
QY 121 ctgctacttcataccatgcctttaaaggggcagtttcttcaaaagcagaacatgcc 180
DB 428 CTGCGTACTTATACCATGCTTAAAGAGGGGCAGTTTCTCAAAAGCAGAAATGCC 369
QY 181 gccagttctcaagtttctctcctaaactccatcttgaatgtaaggggcagctggcccccaatg 240
DB 368 GCCAGTTCTCAAGTTTCTCTTAATCTTGAATGTAAGGGCAGCTGGCCCCCAAG 309
QY 241 tggggagggtcccaaacatttctgaaatcccatcttctgttcgcggctaaatgacagttt 300
DB 309 TGGGAGGTCCGAACTTTCTGAAATCCCATTTCTGTTCGGGCTAAATGACAGATT 249
QY 301 ctgtcattacttagatcccgatctttcccaaaaggtgttgcattacaaagggccagctca 360
DB 248 CTGTCATTACTTAGATT-CCGATCTTTCCTCAAGGTGTTGATTTACAAAGAGGCCAGCTA 150
QY 361 atagccagaatacatgccctgaagagagatgaaatttccaaactgagcagcagga 420
DB 169 ATAG-CAGAAATCATGACCTGTAAGAGAGATGAAA-TTCAAGCTGTGAGCCAGGCAGCA 132
QY 421 gctccagtagtgcgaaggtttcttgagaatcagccattttggtacaaaagagatttttaag 480
DB 131 GCT-CAATATGGCAAGGTCTTTGAGAAATCAGCCATTTGCTACAAAAAGATTTTAAAG 73
QY 481 cttttatgttatccatggagccatgaaaggtcatggtattgtttaagaactatttttaa 540
DB 72 CTTTATGTTATACCATGGAGCCATGAGAAAGGCTATGGATTTGTTTAAGAACTATTTTAAA 13
QY 541 gtgttccagacc 552
DB 12 GTGTTCCAGACC 1
RESULT 4
AI346341/c 549 bp mRNA linear EST 02-FEB-1999
LOCUS qp50b09.x1 NCI_CGAP_C08 Homo sapiens cDNA clone IMAGE:1926425 3'
DEFINITION mRNA sequence.
ACCESSION AI346341
VERSION AI346341.1 GI:4083547
KEYWORDS EST.
SOURCE human.

```

QY 434 aaaggtcttgaatacagccatttggatcacaaaaagatttttaagcttttaagtata 453
|||||
Db 133 AAGAGTCTTGGAGAAATGAGCCATTTGGTACAAAAAGATTTTAAAGCTTTTAAAGTATA 74
|||||

QY 494 ccatggagccatgaaaggctatggatttttaagaactacttttaagtgtttccagacc 553
|||||
Db 73 CCATGGAGCCATGAGAAAGGCTATGGATTGTTTAAAGAACTATTTTAAAGTGTTCAGAGCC 14
|||||

QY 554 aaaaaggaaaaa 566
|||||
Db 13 AAAAAA 1

RESULT 5
AI867454/c
LOCUS AI867454 541 bp mRNA linear EST 17-DEC-1999
DEFINITION w71e10.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2311146 3'.
mRNA sequence.
ACCESSION AI867454
VERSION AI867454.1 GI:5540470
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 541)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1230 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 468.
Location/Qualifiers
1. 541
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2311146"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; Plusmid DNA
from the normalized library NCI_CGAP_GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonoids
1257096-1258631, 1469064-1470983, and 1475552-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 160 a 108 c 110 g 163 t
ORIGIN

Query Match 86.4%; Score 457; DB 9; Length 541;
Best Local Similarity 99.3%; Pred. No. 8.5e-89;
Matches 541; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 16 gatattccataatgattttatctatcgtcatctttacattagccataataatcatttt 75
|||||
Db 541 GATATCCATAATGAATATTTTATATCTGATCCCTTTACATTAGCCATAATACGTTATG 482
|||||

```

```

QY 76 ctgtgatgaagaccctttcacagaatcctatggattgcagcatttctcacttggtacttcata 135
|||||
Db 481 CTGTGATGAGAGACCTTTCACAGAAICCTATGGATTGCAGCAITTCACCTGGCTACITTCATA 422
|||||

QY 136 cccatgcctttaaagagggggcagtttctcaaaagcagaacatgcgcgcagttctcagaatt 195
|||||
Db 421 CCCATGCCITAAAGAGAGGGCGAGTTTCTCAAAAGCAGAAACATGCCGCCAGTTCTCAAGTT 362
|||||

QY 196 ttctctcctaactccatttgaatgtaaggcagctggcccccaatgtggggagggtccgaac 255
|||||
Db 361 TTCTCTCTAACTCCATTTGAATGTAGGGCAGCTGCCGCCCAATGTGGGAGGTCCGAAC 302
|||||

QY 256 atttctgaattcccatcttctctgttcgggggtataatgcagtttctgtcattacttaga 315
|||||
Db 301 ATTTCTGAATTCCTCATTTCTGTTCCGGCTAAATGACAGTTTCTGTCACTACTTAGA 242
|||||

QY 316 ttcccgatcttccccaaagggtgtgtatttaccaaaaggccagcttaataagcagaataatcat 375
|||||
Db 241 TT-CCGATCITTTCCCAAGGTGTTGATTTACAAAGAGGGCAGCTAATAG-CAGAAATCAT 184
|||||

QY 376 gacctgaaagagagagtgaaatttcaagctgtgagccagcagcagcagctcagctatggcaa 435
|||||
Db 183 GACCCGTAAAGAGAGAGATGAAA-TTCAAGCTGTGAGCCAGCGCAGAGCT-CAGTATGGCAA 126
|||||

QY 436 agttctcttgagaatagccatttggtgtacaaaaagattttttaaagcttttatattatacc 495
|||||
Db 125 AGCTTCTTGAGAAATCGCCATTTGGTACAAAAAGATTTTAAAGCTTTTATGTTATACC 66
|||||

QY 496 atggagccatagaaggctatgattgtttaagaactattttaaagtggttccagaccacaa 555
|||||
Db 65 ATGAGCCATAGAAAGCTATGGATTGTTTAAAGTACTATTTTAAAGTGITCCAGACCCAA 6
|||||

QY 556 aaggg 560
|||||
Db 5 AAAGG 1

RESULT 6
BM263376/c
LOCUS BM263376 532 bp mRNA linear EST 18-DEC-2001
DEFINITION Iq29b01.x1 Hke5 Islet Homo sapiens cDNA 3', mRNA sequence.
ACCESSION BM263376
VERSION BM263376.1 GI:17926416
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 532)
Mellon, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
Hillier, L., Barra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,
Schmitt, A., Theisinger, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
, M., Gibbons, H., McCann, R., Cole, R., Tsagaris, V., Williams, T.,
Jackson, Y., and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Mellon, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmellon@biohpc.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Dr. Hiroshi Inoue
(hinoue@im.wustl.edu)
Seq primer: -400P from Gibco
High quality sequence stop: 447.
Location/Qualifiers
1. 532

```

FEATURES
source

Seq primer: M13 Forward	Location/Qualifiers	
POLYA=yes		
1. .454	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:2705051"	
	/clone_lib="NCI_CGAP_Subl"	
	/lab_host="DH10B (Life Technologies)"	
	/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; The NCI_CGAP_Subl library is a subtracted library derived from BI. BI constitutes a mixture of 21 normalized or subtracted NCI_CGAP libraries: NCI_CGAP_C04, NCI_CGAP_Pr22, NCI_CGAP_Pr28, NCI_CGAP_C010, NCI_CGAP_C016, NCI_CGAP_Kid5, NCI_CGAP_Kid12, NCI_CGAP_Kid3, NCI_CGAP_Kid11, NCI_CGAP_Lym2, NCI_CGAP_Pr2, NCI_CGAP_C08, NCI_CGAP_C011, NCI_CGAP_C012, NCI_CGAP_Brn23, NCI_CGAP_Lu5, NCI_CGAP_Lu24, NCI_CGAP_Lu19, NCI_CGAP_C04, NCI_CGAP_C06, NCI_CGAP_Brn25. These 21 libraries were pooled and a single-stranded DNA preparation of the resulting mixture was used as a tracer in a subtractive hybridization with a driver whose composition is detailed below: NCI_CGAP_Kid3 pool 1 LAM 3334-3337, 3682-3683, 3798-3803 --(IMAGE ClonedIDs 132376-132391), 145608-145675, 150052-150055, 1776-3778 (IMAGE ClonedIDs 1323912-1325831, 1471368-1472503, 1492104-1493255) NCI_CGAP_Lu5 pool 1 LAM 3575-3582, 3851-3854 (IMAGE ClonedIDs 1414920-1417951, 1520904-1522439) NCI_CGAP_C04 pool 1 LAM 3164-3167, 3716-3720, 3733-3735 (IMAGE ClonedIDs 1257096-1258631, 1463064-1470983, 1475592-1476743) NCI_CGAP_Pr22 pool 1 LAM 2457-2459, 2758-2759, 3062-3068 (IMAGE ClonedIDs 985608-986759, 1101152-1101959, 1217928-1220615) NCI_CGAP_C010 pool 1 LAM 2644-2653, 2871-2872 (IMAGE ClonedIDs 1057416-1061255, 1144584-1145351) The resulting subtracted library contained 530,000 recombinants. Subtraction was performed as previously described [Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806. TAG_LIB=NCI_CGAP_Brn23 TAG_TISSUE=brain TAG_SEQ=ATATC	
BASE COUNT	126 a 93 c 91 g 144 t	
ORIGIN		
Query Match	69.9%; Score 402; DB 9; Length 454;	
Best Local Similarity	98.0%; Pred. No. 5.6e-70;	
Matches 449; Conservative	0; Mismatches 5; Indels 4; Gaps 4;	
QY	113 gacttcacttgctactcaccatgccttaagaggcgagcttctcaaaagcaga 172	
Db	454 GCATTTCACTTGATCTCATACCAATGCCCTTAAGAGGGCGCATTTCTCAAAAGCAGA 395	
QY	173 aacatgcgcagcttctcactcactccattgaattgaattgaaggcgagctggc 252	
Db	394 AACATGCCCCAGCTTCTCAAGTTTCTCTCACTCACTTCAATTCGATTCGAGGCGACGCTGC 335	
QY	233 ccccaatgtaggagatccgaacattctcgaattcccaattctctgctcgcggtctaaat 292	
Db	334 CCCCAGTGTGGGAGGTCGCAACATTTCTGAATTCCTCCATTTCTTGTTCGCGCTAAT 275	
QY	293 gacagttctctacttagattcccgatcttcccaaaagtgctgatttacaagaag 352	
Db	274 GACAGTTCTGTCTATTACTTAGATT-CCGATCTTTTCCCAAGGTGTGTGATTTCACAAAGAG 216	
QY	353 gcaagctaatagcagaataatcatgacctgaagagagatgaaattcccaagctgtgagcc 412	
Db	215 CCCAGCTAATAG-CAGAAATCATGACCCCTCAAGAGAGAGATGAAA-TTCAAGCTGTGAGCC 158	
QY	413 aggcaggagctccagtagggcaaggtctcttgagaatcagccatttgggtacaaaaaagat 472	
Db	157 AGCAGGAGCT-CAGTATGCAAGAGGTTCTTGACATCAGCCATTTGGTACAAAAAGAT 99	
QY	473 ttttaaaagcttttattatattaccatgagccatagaaagctatgatttcttcaagaact 532	
Db	98 TTTTAAAGCTTTTATGTTATACCAATGAGGACATAGAAAGCTATGTAITGTTTAAAGACT 39	
QY	533 attttaaagtgctccatgaccccaaaagagaaaaaaa 570	
Db	38 ATTTTAAAGTGTTCAGGACCCCAAAAAAATAAAAAA 1	
RESULT 12		
N54784/c		
LOCUS	N54784 486 bp mRNA linear EST 28-JAN-1997	
DEFINITION	yv31a04.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:244302 3', mRNA sequence.	
ACCESSION	N54784	
VERSION	N54784.1 GI:1136104	
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
AUTHORS	1 (bases 1 to 486)	
	Hillier, L., Peterson, G., Becker, M., Bonaldo, M.F., Chispe, B., Chispe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevas, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.	
TITLE	Generation and analysis of 280,000 human expressed sequence tags	
JOURNAL	Genome Res. 6 (9), 807-828 (1996)	
MEDLINE	97044478	
COMMENT	Contact: Wilson RK Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: estewatson.wustl.edu This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert length: 1315 Std Error: 0.00 Seq primer: m13 -40 forward High quality sequence stop: 317. Location/Qualifiers 1. .486 /organism="Homo sapiens" /db_xref="GDB:3793548" /db_xref="taxon:9606" /clone="IMAGE:244302" /clone_lib="Soares fetal liver spleen INFLS" /sex="male" /dev_stage="20 week-post conception fetus" /lab_host="DH10B (ampicillin resistant)" /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Pac 1; Site 2: Eco RI; 1st strand cDNA was primed with a Pac 1 - oligo(dT) primer (5' ACTGGAAGATTAATTAAGATCTTTTITTTTITTTT 3'). Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac 1 and cloned into the Pac 1 and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."	
BASE COUNT	141 a 102 c 139 t	
ORIGIN		
Query Match	69.1%; Score 397.6; DB 10; Length 486;	
Best Local Similarity	97.3%; Pred. No. 4.1e-69;	
Matches 477; Conservative	0; Mismatches 6; Indels 7; Gaps 7;	
QY	67 acgtattgttgatgaagacctttcagaatcctatgagtgagcagcatttcacttgc 126	


```

|||||
Db 486 ACGTATTGCTTGATGAAGACCTTTTACNGAATCCTATGGATTGCGACATTCACATTCGCG 427
Qy 127 tacttcata-cccatgccttaaaagagggcagtttctctcaaaagcagaacacatgccg-cca 184
|||||
Db 426 TACTTCAATCCCATGCTTTAAAGAGGGGCGCACTTCTCAAAGCAGAAACATGCCGCCCA 367
Qy 185 gtctcaagtttctctcaactccattgaatgaatgaaggcagctggcccccgaatgtggg 244
|||||
Db 366 GTTCTCAAGTTTCTCTCACTCCATTTGAATGTAAGGCGAGCTGCCCCCAATGTGGG 307
Qy 245 gaggtccgaaca-ttttctgaattcccatcttcttctgtctcgcgctaaatgacagtttctg 303
|||||
Db 306 GAGGTCCGAACATTTTCTGAATTCCTATTTCTGTGTGCGGCTAAAGACAGATTCTG 247
Qy 304 tcattactagattcccgatctcttcccaagggtgtgatttacaagaagagccagcttaata 363
|||||
Db 246 TCATTACTTAGATT-CCGATCTTTCCCAAGGTTGTGATTTACAAAGAGGCCAGCTAATA 188
Qy 364 gccagaatcatgacctgaaagagagatgaatttcaagctgtgagccagggcagagct 423
|||||
Db 187 G-CAGAATCATGACCTGAAAGAGAGATGAAA-TTCAAGCTGTGACCCAGGCGAGGAGCT 130
Qy 424 ccagtagggcaagggtttcttgagatcagccatttggtagacaaaagagatttttaagctt 483
|||||
Db 129 -CAGTAGGCAAGGTTCTTGAGAAATCAGGCATTTGGTAGCAAAAAGAGATTTTTAAAGCT 71
Qy 484 ttatgttataccatggagccatgaaggctatgattgttttaagaactcttttaagytg 543
|||||
Db 70 TTATGTTATACCATGGAGCCATGAAGGCTATGATGTTTGAAGAACTATTTTAAAGTG 11
Qy 544 ttccagagccc 553
|||||
Db 10 TTCCAGAGCCC 1

RESULT 13
AA775552/c
LOCUS
DEFINITION
IMAGE: 378002 3', mRNA sequence.
ACCESSION
AA775552
VERSION
AA775552.1 GI:2834886
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 432)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wyllie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Contact: Wilson RK
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLNL: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 431.
Location/Qualifiers
1..432
/organism="Homo sapiens"
/db_xref="GDB:1286238"
/db_xref="taxon:9606"
/clone="IMAGE:378002"
/clone_lib="Soares_fetal_heart_NBHL19W"
/sex="unknown"
FEATURES
Source

```

```

/dev_stage="1y weeks"
/lab_host="DHL10B (ampicillin resistant)"
/notes="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not 1; Site_2: Eco RI; 1st
strand cDNA was primed with a Not 1 - oligo(dT) primer [5.
TGTTACCAATCTGAAGTGGAGCGCGCATTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not 1 and cloned into
the Not 1 and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHL19W."
BASE COUNT 122 a 92 c 87 g 131 t
ORIGIN

Query Match. 66.9%; Score 384.8; DB 9; Length 432;
Best Local Similarity 98.6%; Pred. NO. 1.4e-66;
Matches 430; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

Qy 126 ctacttcataccatgccttaaaagagggcagtttctcaaaagcagaacacatgcccgag 185
|||||
Db 432 CTACTTCATACCCATGCCITAAAGAGGGCGAGTTTCTCAAAGCAGAAACATGCCGCCAG 373
Qy 186 ttctcaagtttctctctcaactccatttgaatgaaggcagctggcccccgaatgtgggg 245
|||||
Db 372 TTCGAAGTTTCTCTTAACCTCAATTCGAATGTAAGGCGCAGCTGGCCCATATGTGGGG 313
Qy 246 aggtccgaacatttctgaattcccatcttctgttcgagctaaagcagctttctgctc 305
|||||
Db 312 AGGTCCGAACATTTCTGAATTCCTATTTCTGTGTGCGGCTAAAGACAGTTTCTGTC 253
Qy 306 attacttagattcccgatcttctcccaagggtgtgatttacaagaagggcagctaatagc 365
|||||
Db 252 ATTACTTAGATT-CCGATCTTTCCCAAGGTTGTGATTTACAAAGAGCGCGAGCTAATAG- 193
Qy 366 cagaatcatgacctgaaagagagatgaatttcaagctgtgagccagggcagagctcc 425
|||||
Db 194 CAGAAATCATGACCCCTGAAAGAGAGATGAAA-TTCAAGCTGTGAGCCAGCAGAGAGCT-C 137
Qy 426 agtatggcaagggtctcttgagaatcagccatttggtagacaaaagagatttttaagcttt 485
|||||
Db 136 AGTAGGCAAGGCTTCTTGAGAAATCAGCCATTTGGTAGCAAAAAGAGATTTTAAAGCTTT 77
Qy 486 atgttataccatggagccatgaaggctatgattgttttaagaactatttttaagytgt 545
|||||
Db 76 AIGTTATACCATGGAGCCATAGAAAGGCTATGAGATTGTTTGAAGAACTATTTTAAAGTGT 17
Qy 546 ccagaccaccaaaaagga 561
|||||
Db 16 CCAGACCCCAAAAAGGA 1

RESULT 14
AA626243/c
LOCUS
DEFINITION
IMAGE: 378008 s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:745526
3', mRNA sequence.
ACCESSION
AA626243
VERSION
AA626243.1 GI:2538630
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 410)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wyllie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
TITLE

```



```

VERSION BE047111.1 GI:8364164
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 472)
JOURNAL NCI/NIH-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT National Cancer Institute / National Institute of Dental Research,
Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-femail.nih.gov
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 419.
Location/Qualifiers
FEATURES
source
1..472
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP_HN13"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: tongue; Vector: pCMV-SPO16; Site.1: NciI;
Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.3 kb. Library constructed by Life
Technologies."
BASE COUNT 176 a 70 c 67 g 159 t
ORIGIN
|||||

Query Match 71.0%; Score 419; DB 9; Length 472;
Best Local Similarity 99.8%; Pred. No. 2.7e-50;
Matches 430; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ggcctctgtatctttttactgattgacattggaatgtaattgcgttaataacgc 60
DL 431 GTGCTCTGTATTTTACTGTCATTAGACATTGAATAGTAATTTTCGGTAAAGATACGC 372

QY 61 ttaaggctcttgcaccatgttccctttgtagcaataaaatgttttttaaaact 120
DB 371 TTAAGGCTCTTGTGACCATGTTCCCTTTGTAGCAATAAATGTTTTTACGAAAC 312

QY 121 tctccctggtattgacagtttataatgaacagagttccatcaatgaaatgattttttaa 180
DB 311 TTCTCCCTGGATTAGCAGTTTAAATGAACAGAGTTTCATCAATGAATGAGTATTTAAA 252

QY 181 taaaattgcttaattgctatcagttcagttccacaaatattttaaagtatgctgaaagac 240
DB 251 TAAAATTGCTTAATGATATATCAGTTTCAGTTTCAGTTTCAGTTTCAGTTTCAGTTTC 192

QY 241 ttgaattaaag-aaaaaaaattctcaatcatatttttaaaataagaactaaatttctt 259
DB 191 TTGAATTAAAGAAAAAATAATTTCAATCATATTTTAAATAATAAGACTAAATTTGTT 132

QY 300 tttaaacacattctcaatagaagtgttggaactgaccttatttatactctttttaaag 359
DB 131 TTAAACACATTTTCAATAGAAAGTGAGTTTGAACACTGACCTTATTTATATCTTTTAA 72

QY 360 ttgttctcttcccttgctgctgctgctgctgctgctgctgctgctgctgctgctgct 419
DB 71 TTGTCTCTTTTCCCTGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 12

QY 420 acaagatttctc 430
|||||

```

```

DB 11 ACAAGTIIIC 1

RESULT 4
LOCUS AI983514/c
DEFINITION 233a12 xl NCI-CGAP_Brn53 Homo sapiens cDNA clone IMAGE:2559838 3',
mRNA sequence.
ACCESSION AI983514
VERSION AI983514.1 GI:5810733
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 511)
REFERENCE NCI/NIH-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BIOP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-femail.nih.gov
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life
Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/tbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 423.
Location/Qualifiers
FEATURES
source
1..511
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP_Brn53"
/tissue_type="three pooled meningiomas"
/lab_host="DH10B"
/notes="Organ: Brain; Vector: pCMV-SPO16; Site.1: SalI;
Site.2: NciI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies."
BASE COUNT 150 a 80 c 73 g 168 t
ORIGIN
|||||

Query Match 70.2%; Score 414.4; DB 9; Length 511;
Best Local Similarity 99.5%; Pred. No. 1.2e-49;
Matches 426; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 ggcctctgtatctttttactgctttagcaatgaaatgtaattgcgttaataacgc 60
DB 428 GTGCTCTGTGTAATTTTTTACIGCATTAACATTCGAATAGTAAATTCGGTAAAGATACGC 369

QY 61 ttaaggctcttgcaccatgttccctttgtagcaataaaatgttttttacgaaaaact 120
DB 368 TTAAGGCTCTTGTGACCATGTTCCCTTTGTAGCAATAAATGTTTTTACGAAACT 309

QY 121 tctccctggtattgacagtttataatgaaacagagttccatcaatgaaatgattttttaa 180
DB 308 TTCTCCTCGATTAGTAAATGAACAGAGTTTCATCAATGAATGAGTATTTTAA 249

QY 181 taaaaatttg-cttcaatgctcagttcagctcacaagtatttttaagatgattgagaaga 239
DB 248 TAAAAATTTGCCCTCAATGATCATGTTCAAGTTCACAGTATTTTAAAGATGATGAGA 189

QY 240 cttgaattaaagaagaagaataattctcaatcatatttttaaaataagaactaaattgtt 299
DB 188 CTGAATTAAAGAAAAAATAATTTCAATCATATTTTAAAAAATAAGACTAAATTTGTT 129

QY 300 tttaaacacattctcaatagaagtgttggaactgaccttatttatactctttttaaag 359
|||||

```

[illegible][illegible]

germinal center B cells by flow sorting (CD20⁺, IgD⁺), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGAGCGCGCTCAATTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 151 a 58 c 58 g 144 t
ORIGIN

Query Match 67.6%; Score 399; DB 9; Length 411;
Best Local Similarity 99.8%; Pred. No. 1.9e-47;
Matches 410; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 18 ttactgattagacattgaatagtaatttgcgttaagatagcgtttaaaggtctctttgtga 77
DB 411 TTAAGTGAATGAAACAGAGTTCAATGAATGAGTATTTAAATATAAAATTTGGCTTAAT 232

QY 78 ccattgtttcccttgcgtagcaataaaatgtttttttaaagaaactttctccctgattagca 137
DB 351 CCATGTTTCCCTTTGTAGCAATAAAATCTTTTACGAAAACTTCTCCCTGGATTAGCA 292

QY 138 gtttaaatgaacagagttcatcaatgaatgaatgtattttaaataaaatgtccttaatt 157
DB 291 GTTTAAATGAAACAGAGTTCAATGAATGAGTATTTAAATATAAAATTTGGCTTAAT 232

QY 198 gtatcagttcagctcacaaagtatttttaagatgattggaagacttgtaataag-aaaa 256
DB 231 GTATCAGTTCAAGTTCACAAAGTATTTTAAGATGATTGAGAGACTTGAATTAAGAAAAA 172

QY 257 aaaaatcctcaatcattttttaaataataagactaaaatgtttttttaaacaacatttcaa 316
DB 171 AAAATCTCAATCATATATTTTAAATATAAGACTTAAATGTTTIIIAAAACACATTTCAA 112

QY 317 atagaagtgcgttgcagctgacctattttatactctttttaaagttgttccctttccctg 376
DB 111 ATAGAAGTGAGTTTGAAGTCAAGTCACTGACCTTTTATACICTTTTAAAGTTTGTCTTTCCCTG 52

QY 377 tgcctgtgcataatcttcaagctctgtgcataataacatttgcatacaagtt 427
DB 51 TGCCGTGTCAAACTCTCAAGCTCTGCTGAAATACATTTGATACAAAGTT 1

RESULT 9
AA214632/z
LOCUS 2189h03.s1 NCI_CGAP_GCB1 416 bp mRNA linear EST 13-AUG-1997
DEFINITION mRNA sequence.

ACCESSION AA214632
VERSION AA214632.1 GI:1813269
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 416)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov

COMMENT This clone is available royalty-free through LINL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -41m3 fwd. ET from Amersham
High quality sequence stop: 246.
Location/Qualifiers
1. .416
/organism="Homo sapiens"

FEATURES
source

/db_xref="GDB:5586221"
/db_xref="taxon:9606"
/clone="IMAGE:662901"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="pH108"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20⁺, IgD⁺), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGAGCGCGCTCAATTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 153 a 59 c 60 g 141 t 3 others
ORIGIN

Query Match 66.6%; Score 393; DB 9; Length 416;
Best Local Similarity 97.8%; Pred. No. 1.3e-46;
Matches 407; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 13 ttttttactgcattagacattgaatagtaatttgcgttaagatagcgtttaaaggtcttt 72
DB 416 TTTTITACTGCAATGACCAATGAAATAGTCAATGGCGTAAAGTACGCTTAAAGGCTCTT 357

QY 73 tttgacctatttcccttttgcgttagcaataaaatgtttttttaaagaaactttctccctggat 132
DB 356 TGTGACCATTTTCCCTTTTGTNGCAATAAAATGTTTTTACGAAACTTTCGCCGTGAT 297

QY 133 tagcagtttaaatgaacagagttcctcaatgaatgaatgagttatctaaataaaatctgcc 192
DB 256 TACAGTTTAAATGAAACAGAGTTTCATCAATGAATGAGTATTTAAATATAAAATTTGCC 237

QY 193 ttaagtatcagttcagctcacaaagtatttttaagatgattgagaagacttgaattaaag- 251
DB 236 TTAATGTATCAGTTCACTCAAGTATTTTAAGATGATTCAGAGACTTGAATTAAGA 177

QY 252 aaaaaaaattctcaatcattatttttaaaataaagaactaaatgttttttaaaacaacat 311
DB 176 AAAAAAAATTCNCAATCATATTTTAAAAATATAAGACTAAAAATTTGTTTIAAAACACAT 117

QY 312 ttcaaatagaagtgagtttgaaactgacctatttatactctttttaaagttgttcccttt 371
DB 116 TCAATATAGAGTGAGTTTGAAGTCAAGTATTTTAACTCTTTTAAAGTTTGTCTCTTT 57

QY 372 cctctgtcctgtgtcaaatcttcaagcttcttgcgtgaaataacatttgcatacaagtt 427
DB 56 CCCGTGCTGTGTCAAAATCTTCAAGTCTTGTCTGAAAAATACATTTGATACAAAGTT 1

RESULT 10
BF197806/c
LOCUS 7p92f12.x1 NCI_CGAP_Skl1 Homo sapiens cDNA clone IMAGE:3653518 3', mRNA sequence.
DEFINITION mRNA sequence.

ACCESSION BF197806
VERSION BF197806.1 GI:11087256
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 410)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
JOURNAL

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LINL)
Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL, send email to: info@image.llnl.gov
High quality sequence stop: 404.

FEATURES
source 1..410
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3653518"
/clone_lib="NCI_CGAP_Sknl"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: skin, normal, 4 pooled samples; Vector: PCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.03kb. Library constructed by Life Technologies, cat # 11132-018."

BASE COUNT 148 a 58 c 57 g 147 t
ORIGIN

Query Match 66.4%; Score 391.8; DB 10; Length 410;
Best Local Similarity 99.3%; Pred. No. 1.9e-46;
Matches 404; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 23 gcattagacattgaatgaattgqcttaagatacgccttaaaagctctttgtgaccatg 82
Db 410 GCATTAGACATGAATGATGCTTTCGTTAAGATACGCTTAAGGCTCTTTGACCATG 351
Qy 83 ttccctttgttagcaataaaatgttttttcagaaaactctctccctggattagcagttta 142
Db 350 TTCCCTTTGTPAGCAATAAATGTTTTTTCAGAAAACCTTTCCTCCCTGGATTAGCAGTTA 251
Qy 143 aatgaaacagagttcatcaatgaatgagtgattttaaaataaaatttgccttaattgtatc 202
Db 250 AATGAACACAGATTCATCAATGAATGAGTATTTAAATATAAATTTGCGTTAAIGATC 231
Qy 203 agttcagctcaagaatttttttaagatgattgagaagactcgaaatcaag-aaaaaaat 261
Db 230 AGTTCAGCTCACAAAGTATTTTAAGATGATTGAGAAGACTTGAATTAAGAAAAAATAAAT 171
Qy 262 tctcaatcatatttttaaaacataaagactaaaactgtttttaaaacacatttcaaaataga 321
Db 170 TCTCAATCATATTTTAAATATATAGACTAAAATTTGTTTTAAACACATTTTCAATAGA 111
Qy 322 agtgagttgtaactgaacttatttataactctttttaagttttgtctcttccctgtgacct 381
Db 110 AGTGAGTTTGAATCGACCTATTTATATCTCTTTTAAAGTTTGTTCCTTTTCCCTGTGCTT 51
Qy 382 ggtcaaatcttcaagctctgtctgataaatacatcttgatacaaaagttt 428
Db 50 GGTCAAAATCTCAAGTCTCTGCTGAAAATACATTTTGTATACAAAGTTT 4

RESULT 11
AA150963/c
LOCUS
DEFINITION Z145C01.s1 Soares_pregnant_uterus_NBHPU Homo sapiens cDNA clone
IMAGE:504864 3', mRNA sequence.
ACCESSION AA150963
VERSION AA150963.1 GI:1722474
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 418)

AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Clissold, S., Dietrich, N., DuBuque, J., Faveille, A., Gish, W., Hawkins, M., Hulman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellierberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisakis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R. and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (5), 607-628 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LINL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 985 Std Error: 0.00
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 256.

FEATURES
source 1..418
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="Gene:3809852"
/db_xref="taxon:9606"
/clone="IMAGE:504864"
/clone_lib="Soares_pregnant_uterus_NBHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pRT13-Pac; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACCTGGAAGAAATCGCGGCCCTTTTIIIIIIIIITTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT13 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."

BASE COUNT 151 a 63 c 55 g 139 t 5 others
ORIGIN

Query Match 62.8%; Score 370.4; DB 5; Length 418;
Best Local Similarity 95.7%; Pred. No. 2e-43;
Matches 395; Conservative 0; Mismatches 16; Indels 2; Gaps 2;

Qy 13 ttttttactgtcattgacattgaatgaatttgc-gttaagatcacgtttaaa-ggctc 70
Db 417 TTTTACCHCATTGGGNCATGGAATAGTAAATNGCGGTTAAGATACGCTTAAANGCTC 358
Qy 71 ttltgtaccatgtttccctttgtagcaataaaatgttttttscgaaaaacttctccctg 130
Db 357 TTGTGACCAATGTTCCCTTTGGTAGCAATAAAATGTTTTTTCGAAACACTTTCCTCCG 258
Qy 131 attagcagtttaaatgaacagagttcatcaatgaatgaatgatttttaataataaatttg 190
Db 297 ATTAGCAGTTTAAATGAANCAGAGTTTCATCAATGAATGAGTATTTTAAATAAATAATTG 238
Qy 191 ccttaattgtatcagctcagctcaagatcttttaagatgattgagagactggaattaaa 250
Db 237 CCTTAATGTAATCAGTTCAGCTCAAGATTTTAAATGATGATGAGAGACTTGAATATAA 178
Qy 251 gaaaaaaaatctcaatcatatttttaaaataataagactaaaattgtttttaaacaaca 310
Db 177 GAAAAAAAATTCUACAICATATTTTAAATATATAGACIAAAATTCITTTTAAACACA 118
Qy 311 ttccaataagaagtgtgagttgaaactgaccttatttataactctttttaaagttttccctt 370
Db 117 TTTCAAAATAGAAGTGAAGTTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 58
Qy 371 tccctgtgctgtgctgcaaaatcttcaagctcttgcgtgaaaatacatcttgatcaaaagtt 427
Db 117 TTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 117

with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCCGCTTTTCTTTTCTTTT 3'] , double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p173 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo.

BASE COUNT 129 a 46 c 50 g 135 t
ORIGIN

Query Match 57.6%; Score 340; DB 9; Length 360;
Best Local Similarity 99.7%; Pred. No. 4.2e-39;
Matches 351; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 77 accatgtttccctttgttagcaataaaatgttttttacgaaacttttcccttgattagc 136
DB 360 ACCATGTTCCCTTTGTAGCAATAAAATGTTTTTACGAAACATTCCTCGATTAGC 301
QY 137 agtttaaaagaaacagagtttcatcaatgaatgagttatttaaaataaaatgttcttaa 196
DB 300 AGTTTAAATGAACAGAGTTCATCAATGAATGAATGATTTTAAATTAATAATTTGCTTAA 241
QY 197 tgcattcagttcagctcacagatttttaagatgattgagaagactgaattaaag-aaa 255
DB 240 TGTATCAGTTCAGCTCACAGATTATTTAAGATGATTGAGAGACTTGAATTAAGAAAAA 181
QY 256 aaaaattctcaatcatttttaaaataaagactaaaatgttttaaaacacattcca 315
DB 180 AAAAAATTCATATATTTTAAATAAAGACTAAATGTTTTTAAACACATTTCA 121
QY 316 aatagaagtgcgttgaaactgacctattttatactcttttaagtttgcctttccct 375
DB 120 AATAGAAGTCAGTTGAAGTGAACCTTATTATATCTCTTTTAAAGTTGTTCTCTTCCCT 61
QY 376 gtgcctgtgcataattctcaagctcttgctgaaaaatacatttgacacagtt 427
DB 60 GTGCTGTGTCATAATCTTCAAGTCTTCTGTAATAATACATTGATACAAAGTT 9

RESULT 15
AW023325
LOCUS
DEFINITION
IMAGE:2487196 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

AW023325
467 bp mRNA linear EST 13-SEP-1999
IMAGE:2487196 5', mRNA sequence.
AW023325
GI:5876855
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 467)
Robertson,N.G., Khetarpal,U., Gutierrez-Espeleta,G.A., Bieber,F.R.
and Morton,C.C.
Isolation of novel and known genes from a human fetal cochlear cDNA
library using subtractive hybridization and differential screening
Genomics 23, 42-50 (1994)
95130111
Contact: Morton, C. C.
Departments of Pathology and Obstetrics, Gynecology and
Reproductive Biology
Brigham and Women's Hospital
75 Francis Street, Harvard Medical School, Boston, MA 02115, USA
Tel: 617 732 7980
Fax: 617 738 6996
Email: cmorton@rics.bwh.harvard.edu
DNA sequencing and analyses were performed by National Institutes
of Health Intramural Sequencing Center (NISC; see
http://www.nisc.nih.gov).
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Query Match 59.3%; Score 350; DB 9; Length 370;
Best Local Similarity 99.7%; Pred. No. 1.6e-40;
Matches 361; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 67 gctctttgacacatttccctttgttagcaataaaatgttttttacgaaacttttccct 126
DB 370 GCTCTTTGTGACCATGTTCCCTTTGTAGCAATAAAATGTTTTTACGAAACATTCCTCC 311
QY 127 ctggattagcagtttaaaatgaacagagtttcatcaatgaatgagttatttaaaataaaaa 186
DB 310 CTGGATTAGCAGTTTAAATGAACAGAGTTCATCAATGAATGAATGATTTTAAATAAAAA 251
QY 187 ttgccttaagtatcagttcagctcacagatttttaagatgattgagaagacttgaat 246
DB 250 TTTCCTTTAAATGATCAGTTTCAGTTCACAGATTTTAAAGATGATTGAGAGACTTCAAT 191
QY 247 taaag-aaaaaaattctcaatcatttttaaaataaagactaaaatgtttcttaa 305
DB 190 TAAAGAAAAAATAATCTCAATCATATTTTAAATATAAGACTAAAAATGTTTAAAA 131
QY 306 acacatttcaaatagaagttgagttgaactgacctattttatcaccttttttaagttgtt 365
DB 130 ACACATTTCAATAGAGTGTGTAAGTGAACCTGACCTTATTATATCTCTTTTAAAGTTGTT 71
QY 366 cctttccctgtgcctgtgcataattctcaagctcttgctgaaaaatacatttgatacaaa 425
DB 70 CCTTTTCCCTGTGCTGTGTCATAATCTTCAAGTCTTCTGTAATAATACATTGATACAAAG 11
QY 426 tt 427
DB 10 TT 9

RESULT 14
AA701126/c
LOCUS
DEFINITION
IMAGE:397360 3', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AA701126
360 bp mRNA linear EST 19-DEC-1997
IMAGE:397360 3', mRNA sequence.
AA701126
GI:2704291
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 360)
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisler,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 322.
Location/Qualifiers
1. 360
/organism="Homo sapiens"
/db_xref="GDB:1303262"
/db_xref="taxon:9606"
/clone="IMAGE:397360"
/clone_lib="Soares_pineal_gland_N3HPG"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: pineal gland; Vector: pT713D (Pharmacia)

FEATURES
Source

Tue Apr 30 14:18:20 2002

us-09-248-178-67.std.rst

Plate: L1AM6189 row: G column: 5
Seq primer: M13RPI reverse primer (ABI).
Location/Qualifiers
1. .467
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2487196"
/clone_lib="Morton Fetal Cochlea"
/tissue_type="cochlea"
/dev_stage="16-22 week fetus"
/lab_host="SOLR cells (kanamycin resistant)."
/note="Organ: ear; Vector: pBluescript SK-; Site: 1: EcoRI;
Site_2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned
unidirectionally. Primer: Oligo dT. Fetal cochlea, normal.
37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP
XR Vector. Library constructed by N. Robertson, C. Morton.
-5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTT 3'.
BASE COUNT 161 a 62 c 74 g 170 t
ORIGIN

Query Match 54.1%; Score 319; DB 9; Length 467;
Best Local Similarity 99.4%; Pred. No. 3.5e-36;
Matches 341; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
QY 1 gtgctctgtgtattttttactgcatagacattgaatagtaatttgcgttaagatacgc 60
Db 125 GTGCTCTGTGTATTTTCTTACTGCATTAGACATTCAATAGTAGTAATTCGTTAAGATACGC 184
QY 61 ttaaaggctcttctgaccatgtttcccttctgtagcaataaaatgttttttaccgaaact 120
Db 185 TTAAGGCTCTTCTGACCATGTTCCCTTTGTAGCAATAAAATGTTTTTTACGAAACI 244
QY 121 ttctccctgattagcagtttaaatgaacagagttcatcaatgaatgaatgagatttaaaa 180
Db 245 TTCTCCCTGGATTAGCAGTTTAATCAACACAGAGTTCATCAATGAATGAGTATTAAAA 304
QY 181 taaaaatttgccttaattgtatc-agttcaggtcacaaagtatttttaagatgattgagaaga 239
Db 305 TAAAAATTTCCTTAATGTATCAAGTTCAGCTCACAGTATTTTAAGATGATTGAGAAGA 364
QY 240 ctgaatttaagaaaaaaattctcaatcataattttttaaataat-aagactaaaaattgt 298
Db 365 CTTCGATTTAAGAAAAAAAATTCCTCAATCATATTTTAAATATAAGACATAAAATTGT 424
QY 299 ttttaaacacatttcaaatagagtgagtttgaaactgacctt 341
Db 425 TTTTAAACACATTTCAATAGAGTGAGTTTGAACTGACCTT 467

Search completed: April 29, 2002, 17:17:25
Job time: 13244 sec

Nature Biotechnology, in press. Note that even though the cell type indicated is H1080, since a random activation method was used, these sequence tags are not necessarily expressed in H1080 under normal circumstances."

BASE COUNT 222 a 168 c 139 g 208 t
ORIGIN

Query Match 32.3% Score 301; DB 11; Length 737;
Best Local Similarity 100.0%; Pred. No. 8.1e-83;
Matches 301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 tateccgggttactcgacaaaccatccattccagatggagcgtcccttaccctgcagaa 224
DB 77 TATCCCGGTTACTCGACAAACCCATCACCTTCCAGATGGAGCGCTTACCTGCAGGAA 136
QY 225 taactgttttcaaatatttgggtcttcacacaaacccctatttctgggaagaccctc 284
DB 137 TAACGTGTTTCAATATTGGGCTTTCACCAACCCCTATTTCIGGAAGACCCCTC 196
QY 285 aggtctttaaccccttgagattctccaggggaaattctgaaaaatacatccctatgcct 344
DB 197 AGGCTTTAAACCCCTTCCAGGAAATTTCTGAAAAATACATCCCTATGCCT 256
QY 345 tcataccattctcagctgagattaaggaactgcattgggcagcattttgccataattgagt 404
DB 257 TCATACCATCTCAGCTGGATTAGGAACATGCATTGGGAGCATTTTGCCATAATTGAGT 316
QY 405 gtaagtggcagtggttaactctgcctgccttcagctggctccagctggctccagaccctcaagc 464
DB 317 GTAAAGTGGCAGTGCCATTAACTCTGCTCGCTTCAAGCTGGCTCCAGACCCATCAAGC 376
QY 465 c 465
DB 377 c 377

RESULT 2
BG101072 788 bp mRNA EST 21-APR-2001
LOCUS
DEFINITION R5720398 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG101072
VERSION BG201072.1 GI:13722759
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 788)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R., Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J., Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E., Veloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J. and Ducar,M.
TITLE Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scaina@atersys.com
High quality sequence stop: 514.
Location/Qualifiers
1. .788
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="H1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the

cell type indicated is H1080, since a random activation method was used, these sequence tags are not necessarily expressed in H1080 under normal circumstances."

BASE COUNT 236 a 141 c 146 g 221 t
ORIGIN

Query Match 32.3% Score 301; DB 11; Length 788;
Best Local Similarity 100.0%; Pred. No. 7.6e-83;
Matches 301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 tateccgggttactcgacaaaccatccattccagatggagcgtcccttaccctgcagaa 224
DB 77 TATCCCGGTTACTCGACAAACCCATCACCTTCCAGATGGAGCGCTTACCTGCAGGAA 136
QY 225 taactgttttcaaatatttgggtcttcacacaaacccctatttctgggaagaccctc 284
DB 137 TAACGTGTTTCAATATTGGGCTTTCACCAACCCCTATTTCIGGAAGACCCCTC 196
QY 285 aggtctttaaccccttgagattctccaggggaaattctgaaaaatacatccctatgcct 344
DB 197 AGGCTTTAAACCCCTTCCAGGAAATTTCTGAAAAATACATCCCTATGCCT 256
QY 345 tcataccattctcagctgagattaaggaactgcattgggcagcattttgccataattgagt 404
DB 257 TCATACCATCTCAGCTGGATTAGGAACATGCATTGGGAGCATTTTGCCATAATTGAGT 316
QY 405 gtaagtggcagtggttaactctgcctgccttcagctggctccagctggctccagaccctcaagc 464
DB 317 GTAAAGTGGCAGTGCCATTAACTCTGCTCGCTTCAAGCTGGCTCCAGACCCATCAAGC 376
QY 465 c 465
DB 377 c 377

RESULT 3
BG193485 792 bp mRNA EST 21-APR-2001
LOCUS
DEFINITION R5712619 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG193485
VERSION BG193485.1 GI:13715172
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 792)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R., Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J., Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E., Veloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J. and Ducar,M.
TITLE Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scaina@atersys.com
High quality sequence stop: 553.
Location/Qualifiers
1. .792
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="H1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the

Tue Apr 30 14:18:38 2002

method was used, these sequence tags are not necessarily expressed in H1080 under normal circumstances.

```

BASE COUNT      222 a 190 c 158 g 222 t
ORIGIN

Query Match      31.5%; Score 294; DB 11; Length 792;
Best Local Similarity 99.6%; Pred. No. 1e-80; 1: Indels 1; Gaps 1;
Matches 464; Conservative 0; Mismatches 1;

QY 1 tttactgttgcgaatgacctgagcatcagcagagatgcccagatgaaatcagggaac 60
DB 108 tttactgttgcgaatgacctgagcatcagcagagatgcccagatgaaatcagggaac 167
QY 61 tcttagggatgggtcttctattaccctgggaacacctgagcagatgacctcaccacga 120
DB 168 tcttagggatgggtcttctattaccctgggaacacctgagcagatgacctcaccacga 227
QY 121 tctgcatcaagaatgacctcagcagatgacctcagcagatgacctcagcagatgacctcag 179
DB 228 tctgcatcaagaatgacctcagcagatgacctcagcagatgacctcagcagatgacctcag 287
QY 180 acaaacccatccattccagatgagcagcagcagcagcagcagcagcagcagcagcagcag 239
DB 288 acaaacccatccattccagatgagcagcagcagcagcagcagcagcagcagcagcagcag 347
QY 240 atattgggctcttcacacacacctctattctgggaagacctcagcagcagcagcagcagcag 299
DB 348 atattgggctcttcacacacacctctattctgggaagacctcagcagcagcagcagcagcag 407
QY 300 tggattctccagggaatattctgaaaaatacatccctatgctcctcattcattcagcagcag 359
DB 408 tggattctccagggaatattctgaaaaatacatccctatgctcctcattcattcagcagcag 467
QY 360 ctggattaagaactcattgggagcagcagcagcagcagcagcagcagcagcagcagcagcag 419
DB 468 ctggattaagaactcattgggagcagcagcagcagcagcagcagcagcagcagcagcagcag 527
QY 420 cattaactctgcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 465
DB 528 cattaactctgcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 573

RESULT 4
BG219391 777 bp mRNA EST 21-APR-2001
LOCUS R379149 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION BG219391
ACCESSION BG219391
VERSION BG219391.1 GI:13745412
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 777)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Veloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J.,
and Ducar,M.
Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
Nat. Biotechnol. 19 (5), 440 (2001) In press
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@atersys.com
High quality sequence stop: 529.
Location/Qualifiers
FEATURES
source
1. .777
/organism="Homo sapiens"
/db_xref="Locus:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="H1080"
/note="See 'Creation of Genome-wide Protein Expression'
Libraries using Random Activation of Gene Expression'."

BASE COUNT      230 a 177 c 142 g 226 t
ORIGIN

Query Match      28.8%; Score 269; DB 11; Length 777;
Best Local Similarity 100.0%; Pred. No. 4.6e-73;
Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 ccagatggagcgtccttaccctgaggaataactgcttttcaatatttggcctcttcac 256
DB 111 ccagatggagcgtccttaccctgaggaataactgcttttcaatatttggcctcttcac 170
QY 257 cacaacccctattctgggaagacctcagcagcagcagcagcagcagcagcagcagcagcag 316
DB 171 cacaacccctattctgggaagacctcagcagcagcagcagcagcagcagcagcagcagcag 230
QY 317 aattctgaaaaatacatccctatgcttccatcaccattcagcagcagcagcagcagcagcagcag 376
DB 231 aattctgaaaaatacatccctatgcttccatcaccattcagcagcagcagcagcagcagcagcag 290
QY 377 attgggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 436
DB 291 attgggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 350
QY 437 ttcaagctggctccagacacactcaaggcc 465
DB 351 ttcaagctggctccagacacactcaaggcc 379

RESULT 5
BG218514 787 bp mRNA EST 21-APR-2001
LOCUS R378137 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION BG218514
ACCESSION BG218514
VERSION BG218514.1 GI:13744423
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 787)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Veloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J.,
and Ducar,M.
Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
Nat. Biotechnol. 19 (5), 440 (2001) In press
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@atersys.com
High quality sequence stop: 504.
Location/Qualifiers
FEATURES
source
1. .787
/organism="Homo sapiens"
/db_xref="Locus:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="H1080"
/note="See 'Creation of Genome-wide Protein Expression'
Libraries using Random Activation of Gene Expression'."

```


Nature Biotechnology, in press. Note that even though the cell type indicated is H1080, since a random activation method was used, these sequence tags are not necessarily expressed in H1080 under normal circumstances.

BASE COUNT	236 a	177 c	145 g	228 t	1 others
ORIGIN					
Query Match	26.8%	Score 250;	DB 11;	Length 787;	
Best Local Similarity	59.7%	Pred. No. 2.9e-67;			
Matches 300;	Conservative	0; Mismatches	1;	Indels	0; Gaps
0;					
QY	165	tatcccggttactgacaaacccatccctccagatgacgtccctccatccctgacgaa	224		
DB	77	TATCCCGGTTACTGACAAACCCATCACCTTTCAGATGACGCTCTTACCTGCAGAA	136		
QY	225	taactgtgtttatcaatattgggtctctccaccacacccctatttctggaagaccctc	284		
DB	137	TAACTGTGTTATCAATATTGGGCTCTTACCACACCCCTATTCTGGGAGACCTTC	156		
QY	285	aggtctttaaccccttgagattccaggggaaattctgaaaaatacatccctatgctt	344		
DB	197	AGGTCTTTAACCCCTTGAGATTCCTCAGGGAATTCGAAAAATATACCTCTATGCTT	256		
QY	345	tcataccattctcagctggattaaagaaactgcattgggcagcatttctggaatagtt	404		
DB	257	TCATACCATTCACGTCGATTAAGAACTGCTATTGGCAGCATTTTGCCATAATTGAGT	316		
QY	405	gtaaagtgcagtgccatttaactctccctccatcaagctgctccagaccactcaagc	464		
DB	317	GTAAGTGGCAGTGGCATTAATCTGCTCCGCTTCAAGCTGGCTCCACACCTCAAGGC	376		
QY	465	c 465			
DB	377	c 377			
RESULT	6				
H25624	432 bp	mrna	EST	10-JUL-1995	
LOCUS	Y148g04.r1	Soares breast 3NBHst	Homo sapiens	cdna clone	
DEFINITION	IMAGE:161526 5' similar to gb:J02871 CYTOCHROME P450 1VBI (HUMAN); contains Alu repetitive element; mRNA sequence.				
ACCESSION	H25624				
VERSION	H25624.1	GI:894747			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Woldmann, P. and Wilson, R.				
TITLE	The WashU-Merck EST Project				
JOURNAL	Unpublished (1995)				
COMMENT	Contact: Wilson RK Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Insert Size: 874 High quality sequence stops: 353 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 874 Std Error: 0.00 Seq primer: M13Rpl High quality sequence stop: 353. Location/Qualifiers 1. .432				

BASE COUNT	123 a	98 c	75 g	133 t	3 others
ORIGIN					
Query Match	26.5%	Score 247;	DB 11;	Length 432;	
Best Local Similarity	100.0%	Pred. No. 3.9e-66;			
Matches 247;	Conservative	0; Mismatches	0;	Indels	0; Gaps
0;					
QY	219	caygaataactggtttatcaatatttgggtctctccaccacacccctatttctggaag	278		
DB	85	CAGGAATAACTGTTGTTATCATATATTGGGCTCTTACCACAAACCCCTATTCTGGAAG	144		
QY	279	acctcaggtctttaaaccctctgagattctccagggaataattctgaaaaatacatccct	338		
DB	145	ACCCTCAGGCTCTTAAACCCCTTGAGATTCCTCAGGGAATTCGAAATAATACATCCCT	204		
QY	339	atgcttctataccattctcagctggattaaaggaactgcattgggcagcatttgcataa	398		
DB	205	ATGCTCTATACCATTCACGCTGGATTAAGGAATTCGATTCGCGCAATTTTGCCATAA	264		
QY	399	ttgagtgaatgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag	458		
DB	265	TGAGTGTAAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAG	324		
QY	459	caagccc 465			
DB	325	CAAGGCC 331			
RESULT	7				
A1733538	535 bp	mrna	EST	24-OCT-2000	
LOCUS	Y148g04.y5	Soares breast 3NBHst	Homo sapiens	cdna clone	
DEFINITION	IMAGE:161526 5' similar to SW:CP4B_RABIT P15128 CYTOCHROME P450 4B1				
ACCESSION	A1733538				
VERSION	A1733538.1	GI:5054699			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 535)				
TITLE	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.				
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Other ESTs: y148g04.x5 Contact: Robert Strausberg, Ph.D. Email: cgafts@mail.nih.gov This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. This read is a RESEQUENCE of a previously sequenced human clone Original clone citation: see original entry for original citation information				

/organism="Homo sapiens"
/db_xref="GDB:575683"
/db_xref="taxon:9606"
/clone="IMAGE:161526"
/cdate_lib="Soares breast 3NBHst"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: breast; Vector: pT73D (Pharmacia) with a modified polylinker; Site: 1; Not 1; Site 2: Eco RI; 1st strand cDNA was primed with a Not 1 - oligo(dT) primer [5', TGTTACCAATCTGAGTGGAGCGCGCCCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not 1 and cloned into the Not 1 and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot - 20. Library constructed by Bento Soares and M.Fatima Bonaldo."

This read has been verified (found to hit its original self in the correct orientation)
Insert Length: 874 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 440.

FEATURES

source
1..535
/organism="Homo sapiens"
/db_xref="GDB:575683"
/db_xref="taxon:9606"
/clone="IMAGE:161526"
/clone_lib="Soares breast 3NBHst"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: breast; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGGCCCTTTTITTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 20. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 175 a 109 c 90 g 161 t
ORIGIN

Query Match 26.5%; Score 247; DB 10; Length 535;

Best Local Similarity 100.0%; Pred. No. 3.3e-66; Mismatches 0; Indels 0; Gaps 0;

Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 caggataaactgtttatcaatatttgggctcttcaccacacacccctattttctgggaag 278
DB 85 CAGGAATAACTGTCTTATCAATATTGGGCTCTTCACCACACCCCTATTTCIGGGAAG 144
QY 279 accctcagggttttaaccccttgagattctccagggaataattctgaaaaatacatccct 338
DB 145 ACCCTCAGGTCTTTAAACCCCTTGAGATTCTCCAGGGGAAAATTCGAAAAAATACATCCCT 204
QY 339 atgcttcataccattctcagctggattgaaggaactgcatggcgagcatctttgccataa 396
DB 205 ATGCCCTCATACCATTCACCTGGATTAGGAACATGCAATGGCAGCATTTTGCCTAAT 264
QY 399 ttgagtgtaagtgagtggaacttaactctgtccgcttcagagctgagctccagaccact 458
DB 265 TTGAGTGTAAGTGCGAGTGGCATTAACCTCTGCTCCGCTTCAAGCTGGCTCCAGACCAC 324
QY 459 caagccc 465
DB 325 CAAGGCC 331

RESULT

8
AT820775 548 bp mRNA EST 21-OCT-2000
LOCUS
DEFINITION
Y138c11.y5 Soares breast 3NBHst Homo sapiens cDNA clone
IMAGE:160332.5' similar to SW:CP4B_RABIT P15128 CYTOCHROME P450 451
; mRNA sequence.

ACCESSION
AT820775
VERSION
AT820775.1 GI:5439854
KEYWORDS
EST.

SOURCE

human.
Homo sapiens
Eukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

1 (bases 1 to 548)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index
Unpublished (1997)
Other_ESTs: y138c11.x5

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the IMAGE Consortium (infoimage.lnl.gov) for further information.
This read is a RESEQUENCE of a previously sequenced human clone
Original clone citation: see original entry for original citation
Information

This read has been verified (found to hit its original self in the correct orientation)
Insert Length: 870 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 469.

FEATURES

source
1..548
/organism="Homo sapiens"
/db_xref="GDB:574579"
/db_xref="taxon:9606"
/clone="IMAGE:160532"
/clone_lib="Soares breast 3NBHst"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: breast; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTACCAATCTGAAGTGGAGCGGCCCTTTTITTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 20. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 180 a 111 c 94 g 163 t
ORIGIN

Query Match 26.5%; Score 247; DB 10; Length 548;

Best Local Similarity 100.0%; Pred. No. 3.2e-66; Mismatches 0; Indels 0; Gaps 0;

Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 caggataaactgtttatcaatatttgggctcttcaccacacacccctattttctgggaag 278
DB 99 CAGGAATAACTGTGTGTTATCAATATTGGGCTCTTCACCACACCCCTATTTCIGGGAAG 158
QY 279 accctcagggttttaaccccttgagattctccagggaataattctgaaaaatacatccct 338
DB 159 ACCCTCAGGTCTTTAAACCCCTTGAGATTCTCCAGGGGAAAATTCGAAAAAATACATCCCT 218
QY 339 atgcttcataccattctcagctggattgaaggaactgcatggcgagcatctttgccataa 398
DB 219 ATGCCCTCATACCATTCACCTGGATTAGGAACATGCAATGGCAGCATTTTGCCTAAT 278
QY 399 ttgagtgtaagtgagtggaacttaactctgtccgcttcagagctgagctccagaccact 458
DB 275 TTGAGTGTAAGTGCGAGTGGCATTAACCTCTGCTCCGCTTCAAGCTGGCTCCAGACCAC 338
QY 459 caagccc 465
DB 339 CAAGGCC 345

RESULT

9
BG192933 767 bp mRNA EST 21-APR-2001
LOCUS
DEFINITION
RST12058 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION
BG192933
VERSION
BG192933.1 GI:13714620
KEYWORDS
EST.
SOURCE
human.
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 767)

us-09-248-178-55.rst

Tue Apr 30 14:18:38 2002

JOURNAL.
COMMENT

AUTHORS

Harrington, J. J., Sherf, B., Rundlett, S., Jackson, P. D., Perry, R., Cain, S., Dahl, T., Thornton, M., Kamachandran, R., Whittington, J., Lerner, L., Krashoc, D., McElligott, K., Clark, S., Ways, R., Smith, E., Veloso, N., Jless, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J., and Ducar, M., GeneCoM: a genome-wide protein expression libraries using Random


```

ORIGIN
Db 358 GAGCAATTTTCCTACCAAGGAGAAAGGATAAATAATAACAAAATATATGTAT 259
Qy 624 atggtgtgttgacaaaattatataaacttaggalactctgactcgtgtgttgacatccattaa 683
Db 258 ATGGTGTGTGTGACAAATATATAAATAGGATACTTCGACATCTTCGACATCTTCGACATCTTA 239
Qy 684 cagtaatttttaattcttctgtctgtatctgtgtgaacccacacacacacacacacacacacac 743
Db 236 CAGTAATTTTAAATTCCTGTGTGTATCTGGTGAACCCACAAACACCTTGAAAAAATC 179
Qy 744 aagctga 750
Db 178 AAGCTGA 172

RESULT 17
Bg205754/c 855 bp mRNA EST 21-APR-2001
LOCUS RST25188 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION BG205754
ACCESSION BG205754
VERSION BG205754.1 GI:13727441
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 859)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Veloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J.,
and Ducar,M.
TITLE Creation of Genome-wide Protein Expression Libraries using Random
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scaine@atersys.com
High quality sequence stop: 486.
Location/Qualifiers
1..859
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="H1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is H1080, since a random activation method was used, these sequence tags are not necessarily expressed in H1080 under normal circumstances."

BASE COUNT 248 a 144 c 188 g 279 t
ORIGIN

Query Match 19.1%; Score 178; DB 11; Length 859;
Best Local Similarity 100.0%; Pred. No. 2.6e-45;
Matches 178; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 564 gagacaattttctaccagaagagacacaaagagataataataacaaaatatatgtat 623
Db 361 GAGCAATTTTCCTACCAAGGAGAAAGGATAAATAATAACAAAATATATGTAT 302
Qy 624 atggtgtgttgacaaaattatataaacttaggalactctgactcgtgtgttgacatccattaa 683
Db 301 ATGGTGTGTGTGACAAATATATAAATAGGATACTTCGACATCTTCGACATCTTA 242
Qy 684 cagtaatttttaattcttctgtctgtatctgtgtgaacccacacacacacacacacacacacac 741
Db 241 CAGTAATTTTAAATTCCTGTGTGTATCTGGTGAACCCACAAACACCTTGAAAAAATC 184

RESULT 16
Bg196467/c 882 bp mRNA EST 21-APR-2001
LOCUS RST15689 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION BG196467
ACCESSION BG196467
VERSION BG196467.1 GI:13718154
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 882)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Veloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J.,
and Ducar,M.
TITLE Creation of Genome-wide Protein Expression Libraries using Random
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scaine@atersys.com
High quality sequence stop: 421.
Location/Qualifiers
1..882
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="H1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is H1080, since a random activation method was used, these sequence tags are not necessarily expressed in H1080 under normal circumstances."

BASE COUNT 257 a 144 c 225 g 255 t
ORIGIN

Query Match 20.0%; Score 187; DB 11; Length 882;
Best Local Similarity 100.0%; Pred. No. 4.6e-48;
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 564 gagacaattttctaccagaagagacacaaagagataataataacaaaatatatgtat 623
Db 340 GAGCAATTTTCCTACCAAGGAGAAAGGATAAATAATAACAAAATATATGTAT 261
Qy 624 atggtgtgttgacaaaattatataaacttaggalactctgactcgtgtgttgacatccattaa 683
Db 280 ATGGTGTGTGTGACAAATATATAAATAGGATACTTCGACATCTTCGACATCTTA 221
Qy 684 cagtaatttttaattcttctgtctgtatctgtgtgaacccacacacacacacacacacacacac 743
Db 220 CAGTAATTTTAAATTCCTGTGTGTATCTGGTGAACCCACAAACACCTTGAAAAAATC 161
Qy 744 aagctga 750
Db 160 AAGCTGA 154

RESULT 15
Bg196467/c 882 bp mRNA EST 21-APR-2001
LOCUS RST15689 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION BG196467
ACCESSION BG196467
VERSION BG196467.1 GI:13718154
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 882)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Veloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J.,
and Ducar,M.
TITLE Creation of Genome-wide Protein Expression Libraries using Random
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scaine@atersys.com
High quality sequence stop: 421.
Location/Qualifiers
1..882
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="H1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is H1080, since a random activation method was used, these sequence tags are not necessarily expressed in H1080 under normal circumstances."

BASE COUNT 257 a 144 c 225 g 255 t
ORIGIN

Query Match 20.0%; Score 187; DB 11; Length 882;
Best Local Similarity 100.0%; Pred. No. 4.6e-48;
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 564 gagacaattttctaccagaagagacacaaagagataataataacaaaatatatgtat 623

```

KEYWORDS					
SOURCE Human.					
ORGANISM					
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 					
REFERENCE					
AUTHORS					
TITLE					
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgaps-f@mail.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael K. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio-llnl.gov/bbrp/image/image.html Seq primer: -400P from Gibco High quality sequence stop: 233. Location/Qualifiers					
FEATURES					
source					
1..239					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/clone_image="IMAGE:2751052"					
/clone_lib="NCI-CGAP_Utl"					
/tissue_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"					
/lab_host="DH10B"					
/note="Organ: uterus; Vector: pCMV-SF0K6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.75 kb. Life Technologies catalog #:					
11538-014"					
BASE COUNT 79 a 36 c 54 g 70 t					
ORIGIN					
Query Match 14.8%; Score 138; DB 10; Length 239;					
Best Local Similarity 100.0%; Pred. No. 1.2e+32;					
Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
yy 782 gtaactagtgtagtggtgcttcgaacatgatgtttgatcaaaattccactcagtatctgc 841					
cb 139 gtaactagtgtagtggtgcttcgaacatgatgtttgatcaaaattccactcagtatctgc 80					
yy 842 attaccttttatctcgcaaatatcgcgatgatgttttatctcagtatatctttcccccat 901					
cb 79 attaccttttatctcgcaaatatcgcgatgatgttttatctcagtatatctttcccccat 20					
yy 502 aataaaaaatatctgccca 519					
cb 15 AATAAAAAATATCTGCCA 2					
RESULT 20					
LOCUS H21576					
DEFINITION 332 bp mRNA EST 06-JUL-1995					
Y13c11_r1 Soares breast 3NDHBst Homo sapiens cDNA clone					
IMAGE:fc0532 5' similar to SP:CP4B_RABIT P15128 CYTOCHROME P450					
IVBI ;, mRNA sequence.					
ACCESSION H21576					
VERSION H21576.1 GI:890671					
KEYWORDS EST.					
SOURCE human.					
ORGANISM					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 					
REFERENCE					
AUTHORS					
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman, .M., Hultman,M., Kucaba,T., Le-M., Lennan,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares.M., Tan,F., Trevisakis,E., Waterston					

RESULT 18
 BG212152 807 bp mRNA EST 21-APR-2001
 LOCUS RST31623 Athensys RAGE Library Homo sapiens cDNA, mRNA sequence.
 DEFINITION BG212152
 ACCESSION BG212152.1 GI:13733727
 VERSION EST.
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 807)
 AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,K.,
 Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
 Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
 Veloso,N., Hess,J., Colhoun,K., Lo,K., Offenbacher,J., Danzig,J.,
 and Ducar,M.
 Creation Of Genome-wide Protein Expression Libraries using Random
 Activation Of Gene Expression
 Nat. Biotechnol. 19 (5), 440 (2001) In press
 COMMENT Contact: Scott J. Cain
 Athensys, Inc.
 3201 Carnegie Ave, Cleveland, OH 44115, USA
 Tel: 216 431 9900
 Fax: 216 361 9596
 Email: scain@athensys.com
 High quality sequence stop: 465.
 Location/Qualifiers
 1..807
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Athensys RAGE Library"
 /cell_line="HT1080"
 /note="See 'Creation of Genome-wide Protein Expression
 Libraries using Random Activation of Gene Expression',
 Nature Biotechnology, in press. Note that even though the
 cell type indicated is HT1080, since a random activation
 method was used, these sequence tags are not necessarily
 expressed in HT1080 under normal circumstances."
 BASE COUNT 230 a 200 c 164 g 212 t 1 others
 ORIGIN
 Query Match 18.5%; Score 173; DB 11; Length 807;
 Best Local Similarity 59.6%; Pred. No. 9.3e-44;
 Matches 223; Conservative 0; Mismatches 1; Indels 0; Gaps 0
 QY 242 attgggctctcacacacccctatttctgggaagaccctcaggtcttlaaccccttg 301
 |||||
 DB 284 ATTTGGGCTCTTACCACACCCCTATTCTCTGGGAAGACCCTCAGGCTTTAACCCCTG 343
 QY 302 agattctccgggaataattctgaaaaatacatacctctatgctctcattccatccagct 361
 |||||
 DB 344 AGATTCTCCAGGGAAAAATTCTGAAAAAATACATCCCTATGCTTCATACCATCTAGCT 403
 QY 362 ggattaaggaaactgattggcagcattttgccataatttgagtgtctaaagtggcagtggca 421
 |||||
 DB 404 GGATTAGGAACATGTCATTGGGCAGCATTTTGCCATAATTAGTGTAAAGTGGCAGTGCA 463
 QY 422 ttaactctgctcgcgtctcaagctcgctccagaccactcaagcc 465
 |||||
 DB 464 TTAACCTCTGCCGCTTCAAGCTGGCTCCAGACCACTCAAGGCC 507
 RESULT 19
 AW472756/c
 LOCUS AW472756 239 bp mRNA EST 24-FEB-2000
 DEFINITION xq19a03.x1 NCI_CGAP_U01 Homo sapiens cDNA clone INAGF#2751052 3'.
 mRNA sequence.
 ACCESSION AW472756
 VERSION AW472756.1 GI:7042862


```

/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/notes="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."

BASE COUNT      69 a      58 g      57 t
ORIGIN
Query Match      12.6%; Score 118; DB 11; Length 252;
Best Local Similarity 100.0%; Pred. No. 1.5e-26;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttactgttgcgaagtaccctgagcatcagcagagatgccgagatgaaatcagggaac 60
Dy 107 TTTACTGCTTGGCAAGTACCTGACATCAGCAGATGCCGAGATGAAATCAGGGAC 166

Qy 61 tctaggggatgggtcttattaccctgggaacacctgagccagatgcttacacac 118
Dy 167 TCCTAGGGGATGGTCTTCTATTACCTGGGAACACCTGAGCCAGATGCCCTTACACCAC 224

RESULT 25
BG219912/c      240 bp      mRNA      EST      21-APR-2001
LOCUS
DEFINITION      RST39683 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION      BG219912
VERSION      BG219912.1 GI:13745933
KEYWORDS      EST.
SOURCE      Human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 240)
Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Velooso,N., Hess,J., Cotiren,K., Lo,K., Offenbacher,J., Danzig,J.
and Ducar,M.
Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
Nat. Biotechnol. 19 (5), 440 (2001) In press
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9500
Fax: 216 361 9596
Email: scain@atersys.com
High quality sequence stop: 190.
Location/Qualifiers
1..240
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/notes="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."

BASE COUNT      85 a      40 c      53 g      62 t
ORIGIN
Query Match      11.6%; Score 108; DB 11; Length 240;
Best Local Similarity 100.0%; Pred. No. 1.8e-23;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 85 a      40 c      53 g      62 t
Dy 107 TTTACTGCTTGGCAAGTACCTGACATCAGCAGATGCCGAGATGAAATCAGGGAC 166

Qy 148 acgcaccggtagtaaac 164
Dy 147 ACCGACCCGGTAGTAAAC 163

RESULT 27
AQ505963      729 bp      DNA      GSS      29-APR-1999
LOCUS
DEFINITION      RPCI-11-316A11.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-316A11
DNA sequence.
ACCESSION      AQ505963
VERSION      AQ505963.1 GI:4710710
KEYWORDS      GSS.

```


us-09-248-178-55.rst

Tue Apr 30 14:18:38 2002

ORIGIN

Query Match 2.8%; Score 26; DB 11; Length 483;
 Best Local Similarity 100.0%; Pred. No. 1:le-02;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 200 gatggagcgtcttaccctgcaggaaat 225
 DB 298 GATGGAGCGCTCTTACCTGCAGGAAT 323

RESULT 32
 LOCUS AW465396 561 bp mRNA EST 24-FEB-2000
 DEFINITION BP230019A10E10 Soares normalized bovine placenta Bos taurus cDNA
 ACCESSION AW465396
 VERSION AW465396.1 GI:7035564
 KEYWORDS EST.
 SOURCE COW.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
 Bovidae; Bovinae; Bos.
 REFERENCE 1 (bases 1 to 561)
 AUTHORS Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinas,J., Liu,L. and Larson
 J.H.

TITLE Bovine ESTs
 JOURNAL Unpublished (2000)
 COMMENT Contact: Lewin, H. A.
 W. M. Keck Center for Comparative and Functional Genomics
 University of Illinois at Urbana-Champaign
 340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
 61801, USA
 Tel: 217 333 5998
 Fax: 217 244 5617
 Email: h-lewin@uiuc.edu
 Funding for cattle EST sequencing was provided by the USDA National
 Research Initiative, Animal Genome Resource Grant AG 95-3205-8534
 to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED
 from Washington University Genome Center. Vector Trimmi g:
 Cross-match from Washington University Genome Center PHRAP suite.
 Sequences submitted are vector free and at least 200 bp in length.
 PCR Primers
 FORWARD: TAATACCGCTCACTATAGG
 BACKWARD: ATTAACCGCTCACTAAAG
 Insert Length: 561 Std Error: 0.00
 Plate: BP230019A10 row: E column: 10
 Seq primer: ACCGGATAACAAATTCACACAGGA
 High quality sequence stop: 561.
 Location/Qualifiers
 1..561
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone="BP230019A10E10"
 /clone_lib="Soares normalized bovine placenta"
 /sex="female"
 /lab_host="DH10B"
 /note="Organ: placenta; Vector: pT73pac; Site: 1: EcoRI;
 Site: 2: NotI; The cDNA library was contributed by the
 Soares laboratory and it was constructed and normalized
 as described by Bonaldo, M.F., Lennon, G. and Soares,
 M.B. (1996), Genome Research 6(9): 791-806."

BASE COUNT 119 a 164 c 124 g 154 t

ORIGIN

Query Match 2.8%; Score 26; DB 10; Length 561;
 Best Local Similarity 100.0%; Pred. No. 99;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 200 gatggagcgtcttaccctgcaggaaat 225
 DB 15 GATGGAGCGCTCTTACCTGCAGGAAT 44

RESULT 33
 LOCUS BF043014 572 bp mRNA EST 10-OCT-2000
 DEFINITION BP250009B10D11 Soares normalized bovine placenta Bos taurus cDNA
 ACCESSION BF043014
 VERSION BF043014.1 GI:10760069
 KEYWORDS EST.
 SOURCE COW.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
 Bovidae; Bovinae; Bos.
 REFERENCE 1 (bases 1 to 572)
 AUTHORS Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinas,J., Liu,L. and Larson
 J.H.

TITLE Bovine ESTs
 JOURNAL Unpublished (2000)
 COMMENT Contact: Lewin, H. A.
 W. M. Keck Center for Comparative and Functional Genomics
 University of Illinois at Urbana-Champaign
 340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
 61801, USA
 Tel: 217 333 5998
 Fax: 217 244 5617
 Email: h-lewin@uiuc.edu
 Funding for cattle EST sequencing was provided by the USDA National
 Research Initiative, Animal Genome Resource Grant AG 95-3205-8534
 to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED
 from Washington University Genome Center. Vector Trimmi g:
 Cross-match from Washington University Genome Center PHRAP suite.
 This sequence is vector free and at least 200 bp in length.
 PCR Primers
 FORWARD: TAATACCGCTCACTATAGG
 BACKWARD: ATTAACCGCTCACTAAAG
 Insert Length: 572 Std Error: 0.00
 Plate: BP250009B10 row: D column: 11
 Seq primer: ACCGGATAACAAATTCACACAGGA
 High quality sequence stop: 572.
 Location/Qualifiers
 1..572
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone="BP250009B10D11"
 /clone_lib="Soares normalized bovine placenta"
 /sex="female"
 /lab_host="DH10B"
 /note="Organ: placenta; Vector: pT73pac; Site: 1: EcoRI;
 Site: 2: NotI; The cDNA library was contributed by the
 Soares laboratory and it was constructed and normalized
 as described by Bonaldo, M.F., Lennon, G. and Soares,
 M.B. (1996), Genome Research 6(9): 791-806."

BASE COUNT 139 a 167 c 151 g 115 t

Query Match 2.8%; Score 26; DB 11; Length 572;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 200 gatggagcgtcttaccctgcaggaaat 225